Task-based modeling and optimization of a cone-beam CT scanner for musculoskeletal imaging

P. Prakash and W. Zbijewski
Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland 21205

G. J. Gang
Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland 21205 and Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Ontario M5G 2 M9, Canada

Y. Ding and J. W. Stayman
Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland 21205

J. Yorkston
Carestream Health, Rochester, New York 14615

J. A. Carrino
Department of Radiology, Johns Hopkins University, Baltimore, Maryland 21287

J. H. Siewerdsen[a)]
Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland 21205; Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Ontario M5G 2 M9, Canada; and Department of Radiology, Johns Hopkins University, Baltimore, Maryland 21287

(Received 4 May 2011; revised 27 July 2011; accepted for publication 15 August 2011; published 22 September 2011)

Purpose: This work applies a cascaded systems model for cone-beam CT imaging performance to the design and optimization of a system for musculoskeletal extremity imaging. The model provides a quantitative guide to the selection of system geometry, source and detector components, acquisition techniques, and reconstruction parameters.
Methods: The model is based on cascaded systems analysis of the 3D noise-power spectrum (NPS) and noise-equivalent quanta (NEQ) combined with factors of system geometry (magnification, focal spot size, and scatter-to-primary ratio) and anatomical background clutter. The model was extended to task-based analysis of detectability index (d') for tasks ranging in contrast and frequency content, and d' was computed as a function of system magnification, detector pixel size, focal spot size, kVp, dose, electronic noise, voxel size, and reconstruction filter to examine trade-offs and optima among such factors in multivariate analysis. The model was tested quantitatively versus the measured NPS and qualitatively in cadaver images as a function of kVp, dose, pixel size, and reconstruction filter under conditions corresponding to the proposed scanner.
Results: The analysis quantified trade-offs among factors of spatial resolution, noise, and dose. System magnification (M) was a critical design parameter with strong effect on spatial resolution, dose, and x-ray scatter, and a fairly robust optimum was identified at $M \sim 1.3$ for the imaging tasks considered. The results suggested kVp selection in the range of ~65–90 kVp, the lower end (65 kVp) maximizing subject contrast and the upper end maximizing NEQ (90 kVp). The analysis quantified fairly intuitive results—e.g., ~0.1–0.2 mm pixel size (and a sharp reconstruction filter) optimal for high-frequency tasks (bone detail) compared to ~0.4 mm pixel size (and a smooth reconstruction filter) for low-frequency (soft-tissue) tasks. This result suggests a specific protocol for 1×1 (full-resolution) projection data acquisition followed by full-resolution reconstruction with a sharp filter for high-frequency tasks along with 2×2 binning reconstruction with a smooth filter for low-frequency tasks. The analysis guided selection of specific source and detector components implemented on the proposed scanner. The analysis also quantified the potential benefits and points of diminishing return in focal spot size, reduced electronic noise, finer detector pixels, and low-dose limits of detectability. Theoretical results agreed quantitatively with the measured NPS and qualitatively with evaluation of cadaver images by a musculoskeletal radiologist.
Conclusions: A fairly comprehensive model for 3D imaging performance in cone-beam CT combines factors of quantum noise, system geometry, anatomical background, and imaging task. The analysis provided a valuable, quantitative guide to design, optimization, and technique selection for a musculoskeletal extremities imaging system under development. [DOI: 10.1118/1.3633937]

Key words: musculoskeletal radiology, cone-beam computed tomography, cascaded systems analysis, flat-panel detector, imaging performance, noise-power spectrum, noise-equivalent quanta, anatomical clutter, system optimization, observer performance, detectability index

I. INTRODUCTION

Cone-beam computed tomography (CBCT) with flat-panel detectors (FPDs) has entered a host of applications in preclinical imaging (e.g., small animal scanners), diagnostic imaging (e.g., breast and maxillofacial scanners), and image-guided procedures (e.g., C-arms and medical linear accelerators). The technology provides volumetric imaging from a single rotation about the subject with submillimeter spatial resolution and soft-tissue visibility. As described previously,[1,2] a FPD-CBCT system is being developed for musculoskeletal extremities imaging. The system offers a compact embodiment for imaging the upper or lower extremities that complements CT and MR in the ability to image weight-bearing extremities and the capability for integrated radiography, motion/kinematic imaging, and volumetric imaging within the same platform.

Cascaded systems analysis has been shown to describe the imaging performance of CBCT in terms of Fourier domain metrics, including the fully 3D noise-power spectrum (NPS), modulation transfer function (MTF), and noise-equivalent quanta (NEQ).[3–8] Such analysis can be extended to simple models of observer performance in terms of task-based detectability index (d'),[9] which has been shown to demonstrate reasonable correspondence with real observer performance for a variety of simple imaging tasks across a broad range of experimental conditions.[10]

A considerable body of previous work established the basis for modeling of CBCT imaging performance. For example, Siewerdsen and Jaffray[11] showed a cascaded systems model for 3D NPS and NEQ in CBCT to agree with measurements and to reduce to the classic descriptions of voxel noise. They also exercised the model to examine trade-offs in spatial resolution and noise in the presence of x-ray scatter[12] and the effect of direct- and indirect-detection FPD characteristics on CBCT image quality.[13] Tward and Siewerdsen [5,6] extended the model to task-based detectability index and described the effects of 3D noise aliasing. Such analysis applies to both limited-angle tomography (tomosynthesis)[14,15] and CBCT in that the extent of the source-detector orbit is a free parameter. Factors beyond quantum noise are known to have significant effect on CBCT image quality and can be similarly modeled—most notably, x-ray scatter, focal spot blur, and anatomical background clutter. X-ray scatter and focal spot blur depend strongly on system geometry and have been included in descriptions of the "system" NEQ.[7,16–19] Jain *et al.*[17] and Kyprianou *et al.*[18] extended the analysis of MTF, NPS, and NEQ to include both scatter and focal spot blur, and Kyprianou *et al.*[20] employed the model to investigate the performance of a neurovascular angiography system. The effect of anatomical background on detectability is a subject of considerable interest, commonly modeling the background as power-law noise κ/f^{β} included as an additive noise term in the "generalized" NEQ. For example, in breast imaging, the power-law characteristic in 2D mammographic images was characterized by Bochud *et al.*,[21] Burgess,[22,23] and others[24] and extended to 3D breast imaging by Metheany *et al.*,[25] Engstrom *et al.*,[26] Glick *et al.*,[27] Gong *et al.*,[28] and Reiser and Nishikawa,[29] showing β differs in 2D projections versus 3D tomosynthesis and CBCT reconstructions. In a fairly general context, Gang *et al.*[30] showed power-law background noise to arise from superposition of self-similar (fractal) structures in 2D and 3D images, derived relations between β in projection, tomosynthesis, and CBCT, and incorporated the result in a generalized detectability index. Yoon *et al.*[14] used such modeling to characterize the rejection of background clutter in tomosynthesis, and Siewerdsen *et al.*[31] showed the dependence of CBCT detectability on the clutter factor (β) and the size of the stimulus (e.g., low- or high-frequency tasks).[23,32]

The work below applies such approaches in a model that includes basic cascaded systems analysis parameters (quantum noise), system parameters (x-ray scatter and focal spot blur), and generalized parameters (anatomical background) toward the design and optimization of a CBCT system being developed for musculoskeletal extremities imaging. The research is distinct from previous work not only in the specific area of clinical application but also in the fairly comprehensive model that combines descriptions of quantum, system, and generalized noise factors. The sections below: (1) summarize the cascaded systems model; (2) extend to task-based detectability index; (3) compare to quantitative measurements of NPS and qualitative evaluation of cadaver images; and (4) evaluate the dependence of detectability on system geometry, acquisition techniques, and reconstruction techniques as a guide to design and optimization of an extremities CBCT scanner.

II. THEORETICAL METHODS

II.A. Cascaded systems analysis

The fundamental aspects of cascaded systems analysis have been discussed in previous work[3,4,33,34] in modeling the imaging chain as a series of amplification, blur, and sampling stages, where output from one stage serves as the input for the next. Early work involved application of cascaded systems analysis in modeling of indirect and direct-detection FPDs,[3,4] extended to descriptions of mammography[35] and angiography,[7,19] dual-energy imaging,[36,37] tomosynthesis,[15] and CBCT.[5,30] Variations include incorporation of various system factors (viz., magnification, focal spot size, and x-ray scatter)[16–18] and generalization to include noise factors other than quantum noise (viz., anatomical background clutter).[30,36] Such models have shown good agreement with measured MTF, NPS, detective quantum efficiency (DQE), and NEQ, and have been incorporated in task-based models of detectability index (d') in a manner that yields reasonable correspondence with human observer performance for a variety of simple imaging tasks across a broad range of imaging conditions.[10] A conceptual diagram combining factors of quantum noise, system parameters, background noise, and interdependencies in the NEQ is shown in Fig. 1.

II.B. The 3D NEQ

The NEQ is an important figure of merit describing the number of photons at each spatial frequency for which an

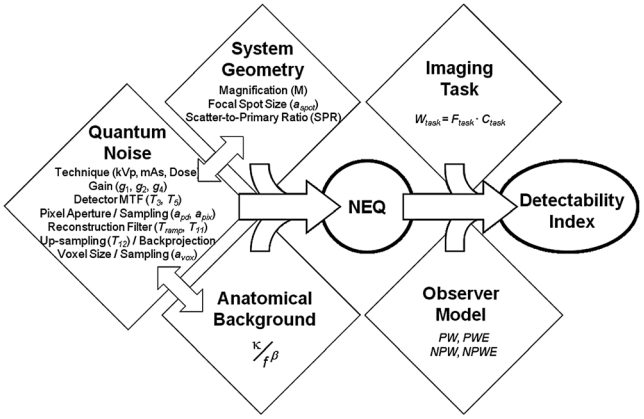

FIG. 1. Flowchart illustration of the model for CBCT imaging performance combining factors of quantum noise, system parameters (focal spot blur, magnification, and x-ray scatter), and background anatomical noise.

ideal detector would produce the same NPS.[38,39] The following sections summarize three distinct forms of NEQ: (1) a form representing only the quantum noise (including factors of kVp, dose, scintillator, FPD design, readout electronics, etc.) propagated through the 2D image acquisition and 3D reconstruction process[5]; (2) an extension to include various system parameters (viz., magnification, focal spot size, and x-ray scatter)[17,18]; and (3) an extension to include additional factors within the denominator of the NEQ (viz., anatomical background noise, originally termed a "generalized" NEQ by Barrett *et al.*[40,41]).[30] Finally, these forms are pulled together in a fairly comprehensive model of 3D NEQ as illustrated in Fig. 1, presenting a broad space of parameters and interdependencies to be considered in imaging system optimization. Details of the individual models and parameters are found in the references noted above and briefly summarized below.

II.B.1. "Quantum" NEQ

As described by Cunningham,[38] there are several equivalent ways to define the NEQ (conceptual, descriptive, stochastic, and predictive). The "stochastic" form can be readily adapted to describe the 3D quantum noise in CBCT[5]

$$NEQ_{\text{quantum}}(f_x,f_y,f_z) = \pi f \frac{T_3^2(f,f_z)T_5^2(f,f_z)T_{11}^2(f)T_{12}^2(f,f_z)}{S_Q^{\text{tot}}(f_x,f_y,f_z) + S_E^{\text{tot}}(f_x,f_y,f_z)}, \quad (1)$$

where f_x, f_y, f_z are the 3D spatial-frequency coordinates, and $f = \sqrt{f_x^2 + f_y^2}$ represents the cylindrically symmetric radial frequency. S_Q^{tot} and S_E^{tot} describe quantum noise and electronics noise associated with the total input quanta, $\overline{q_0^{\text{tot}}}$, respectively (with the superscript tot distinguishing the total fluence from primary and scatter fluence terms, below). The T_i represent the transfer function (MTF) associated with different stages in the imaging chain—e.g., T_3 (scintillator MTF), T_5 (photodiode aperture MTF), T_{11} (apodization filter MTF), and T_{12} (interpolation filter MTF). Notation throughout is as consistent as possible with previous work[5,6] and summarized in Table I. Note that the equation involves fully 3D MTF and NPS terms [the latter with units $(\mu^2)(\text{mm}^3)$, where μ is attenuation coefficient, e.g., (mm^{-1})], giving 3D NEQ with units $[(\text{mm}^{-2})$, interpreted as $(\text{photons}/\text{mm}^2)]$.

TABLE I. Summary of terms used in cascaded systems analysis.

Term	Definition	Nominal value
f_u, f_v	Spatial-frequency domain at the detector plane	—
f_x, f_y, f_z	Spatial-frequency domain at the 3D object volume	—
g_j	Gain at stage j	—
T_j	Transfer function at stage j	—
S_Q	NPS due to quantum noise	—
S_E	NPS due to additive electronics noise	—
κ	Anatomical clutter magnitude	$3.3 \times 10^{-10}\ \mu^2\ \text{mm}^3$
β	Anatomical clutter power-law exponent	3
S_B	Power spectrum due to anatomical background clutter	—
E	Eye-filter transfer function	—
N_i	NPS for observer internal noise	—
$\overline{q_0^{\text{tot}}}$	Mean fluence per projection incident on the detector after attenuation by object	$3.94 \times 10^3\ \text{mm}^{-2}$
X	Exposure per projection at the detector after object attenuation	0.027 mR
a_{pd}	Photodiode width (square)	0.16 mm
a_{pix}	Detector pixel aperture width (square)	0.194 mm
a_{vox}	Voxel size (natural voxel size: $a_{\text{vox}} = a_{\text{pix}}/M$)	0.15 mm
a_{spot}	Focal spot size	0.5 mm
SDD	Source-detector distance	525 mm
SAD	Source-axis distance	410 mm
M	System magnification, $M = SDD/SAD$	1.28
m	Number of projections	480
SPR	Scatter-to-primary ratio	~1.0
σ_{add}	Additive electronics noise	2000 e/pixel
kVp	Beam energy (peak kilovoltage)	90 kVp
D	Dose to water at isocenter (16 cm cylinder)	5 mGy
h_{win}	(Cosine) reconstruction filter parameter	0.5 (Hann)

II.B.2. System NEQ

Factors of magnification, focal spot size, and x-ray scatter have been incorporated within the NEQ under the moniker of "system NEQ" (sometimes called "generalized NEQ," but the former term is used here to avoid confusion with inclusion of anatomical background noise, below, with deference to Barrett's original use of the term "generalized"[40,41]). The x-ray focal spot is a source of blur[16–18] for systems with geometric magnification, M, degrading the numerator of the NEQ (the system MTF) but not the stochastic quantum noise in the denominator, S_Q. X-ray scatter from the object can be similarly included in the system NEQ,[17,18] and introduces blur, noise, and contrast degradation. The x-ray fluence at

the detector includes both primary and scattered radiation (q_0^p and q_0^s, respectively), giving for the system NEQ

$$NEQ_{\text{sys}}(f_x,f_y,f_z) = \pi f \frac{T_3^2 T_5^2 T_{11}^2 T_{12}^2 T_{\text{spot}}^2 T_{\text{scatter}}^2}{S_Q^{p+s} + S_E^{p+s}}, \quad (2)$$

where frequency notation on MTF and NPS terms is implicit but has been dropped for conciseness. T_{spot} is the focal spot MTF, T_{scatter} is the scatter MTF, S_Q^{p+s} is the total quantum noise due to primary and scattered fluence, and S_E^{p+s} is the electronics noise (relative to the primary + scatter fluence as detailed in Appendix C). The superscript on NPS terms S_Q^{p+s} and S_E^{p+s} is distinct from S_Q^{tot} in that it allows for separate consideration of primary and scattered quanta in the 3D NPS (e.g., different x-ray spectra; see Appendix C). All of these terms carry implicit dependence on the system magnification, M. The focal spot was modeled as a uniform rect of width a_{spot}. The scatter MTF was described by Seibert and Boone[42] in the form of a "scatter psf," modeled as the spatial domain product of Gaussian kernel and $1/r$, where r is the radial distance from the origin. The corresponding Fourier domain scatter MTF is a product of a Gaussian and a zeroth-order modified Bessel function with a constant scale factor related to scatter fraction. Kyprianou *et al.*[18] extended the analysis to a composite frequency distribution of blur given by addition of weighted MTF components of both primary and scattered radiations.

II.B.3. Generalized NEQ

The generalized NEQ includes any number of additional noise terms added to the denominator as suggested by Barrett.[40,41] Specifically, anatomical background clutter (i.e., anatomical structures present in the image but not associated with the signal to be detected) is included as an additional power spectrum, $S_B(f, f_z)$. Extending the 3D quantum NEQ of Eq. (1) to include anatomical noise gives

$$NEQ_{\text{gen}}(f_x,f_y,f_z) = \pi f \frac{T_3^2 T_5^2 T_{11}^2 T_{12}^2}{S_Q^{\text{tot}} + S_E^{\text{tot}} + T_3^2 T_5^2 T_{11}^2 T_{12}^2 S_B}, \quad (3)$$

where the various MTF terms are applied to S_B to emphasize that S_B is a property of the object being imaged, and that its presentation within the image is subject to blur by the imaging system. The power spectrum associated with anatomical clutter has been investigated in depth in the context of 2D projection imaging,[21,23,36] and more recently in 3D imaging,[25,29,43] with an empirical power-law (κ/f^β) form being fairly ubiquitous and shown by Gang *et al.*[30] to correspond to the projection (or 3D reconstruction) of a collection of randomly sized and oriented self-similar (fractal) objects. Details of the power-law form and dependencies on other system factors (e.g., beam energy) are discussed in Appendix D.

II.B.4. A comprehensive model

The quantum, system, and generalized forms of the NEQ summarized in Eqs. (1)–(3) can be brought together in a comprehensive form

$$NEQ(f_x,f_y,f_z) = \pi f \frac{T_3^2 T_5^2 T_{11}^2 T_{12}^2 T_{\text{spot}}^2 T_{\text{scatter}}^2}{S_Q^{p+s} + S_E^{p+s} + T_3^2 T_5^2 T_{11}^2 T_{12}^2 T_{\text{spot}}^2 T_{\text{scatter}}^2 S_B}, \quad (4)$$

where Eq. (2) was combined with Eq. (3) to yield a fairly comprehensive model for NEQ that accounts for all three factors: quantum noise, system factors (such as magnification, focal spot size, and x-ray scatter), and background anatomical noise. That is, the system NEQ of Eq. (2) ignores factors explicit in the generalized NEQ of Eq. (3)—and vice versa; however, each can be combined in the fairly comprehensive form of Eq. (4), as each is grounded in the same underlying framework of linear systems analysis. While there is a temptation to continue applying a string of monikers and sub/superscripts (e.g., "generalized system NEQ"), we refer to it simply as the NEQ and refer instead to specific forms in Eqs. (1)–(3) by distinct labels of quantum, system, and generalized, respectively. For the sake of conciseness within the main text, details of individual terms in the NEQ are included in the Appendices. Appendix A describes the MTF terms. Appendix B describes the x-ray scatter model. Appendix C describes the NPS for quantum and electronics noise, S_q and S_E, and Appendix D describes the power-law model of anatomical clutter.

Two simplifying assumptions are made regarding x-ray scatter. First, we assume that x-ray scatter imparts a reduction in contrast, an additive quantum noise source, but does not blur the image—i.e., T_{scatter}^2 is constant, independent of spatial-frequency, given by $1/(1+SPR)$. Analysis without such assumption has been detailed by Kyprianou *et al.*[18] We further assume the x-ray scatter spectrum has the same shape as the primary. The effects neglected in such analysis are believed to be small (e.g., maximum energy shift due to x-ray scatter <10 keV and scatter MTF dominated by focal spot and detector MTF). While neither is a necessary assumption, they allow simplification of the model to a relatively compact analytical form with respect to a system parameter space that is already quite broad. These simplifying assumptions yield a concise form for the NEQ including both system parameters of magnification, focal spot blur, and SPR as well as generalized parameters of anatomical background

$$NEQ(f_x,f_y,f_z) = \frac{\pi f}{(1+SPR)} \cdot \frac{T_3^2 T_5^2 T_{11}^2 T_{12}^2 T_{\text{spot}}^2}{S_Q^{\text{tot}} + S_E^{\text{tot}} + T_3^2 T_5^2 T_{11}^2 T_{12}^2 T_{\text{spot}}^2 \left(\frac{1}{1+SPR}\right) S_B}. \quad (5)$$

The SPR term arises simply from the assumption of a constant T_{scatter} and reproduces the familiar factor $(1/1+SPR)$ as the (low-frequency) degradation due to a scatter. Neglecting the energy shifts in the scatter spectrum (as justified in Appendix C) reduces $S_Q^{p+s}(S_E^{p+s})$ to $S_Q^{\text{tot}}(S_E^{\text{tot}})$. Note that the S_B term [see Eq. (D2) in Appendix D] also depends on SPR in such a way that scatter reduces the contrast of background clutter by a

factor of $1/(1+SPR)$, reducing the third term in the denominator in proportion to $1/(1+SPR)^2$.

II.C. The 3D detectability index

Extension of the NEQ to task-based analysis of detectability index[9] (d') allows investigation of the broad space of system design parameters, acquisition techniques, and 3D reconstruction parameters in terms of a summary objective function (figure of merit) for multivariate optimization in the design of a CBCT scanner. The detectability index has been used as an objective function for system optimization in a variety of applications[30,37,44] and Gang *et al.*[10] have demonstrated reasonable correspondence with real observer performance for simple imaging tasks. The work of Gang *et al.*[10] formed an important theoretical/experimental justification for the current work and provided confidence that the trends and optima computed in this paper will bear correspondence with real observer performance for simple imaging tasks. Basic observer models for d' are as follows:

II.C.1. 3D prewhitening (PW) model

The PW observer simply weights the NEQ by a template (a matched filter or "task function") representing the spatial frequencies of interest in performing the task. It assumes an ideal observer capable of "whitening" (decorrelating) noise in the image[9]

$$d'^2_{PW} = \iiint \frac{MTF^2 W^2_{\text{task}}}{S_Q + S_E + MTF^2 S_B} df_x df_y df_z, \tag{6}$$

where MTF represents the system MTF, $MTF = T_3 T_5 T_{11} T_{12} T_{\text{spot}} T_{\text{scatter}}$, and the integral is performed over the 3D Nyquist frequency region. W_{task} is the frequency domain representation of the imaging task defined as the Fourier transform of binary hypotheses as detailed below.

II.C.2. 3D nonprewhitening (NPW) model

The NPW observer[45] does not decorrelate the image noise but instead applies a detection template in the form of the expected signal

$$d'^2_{NPW} = \frac{\left[\iiint MTF^2 W^2_{\text{task}} df_x df_y df_z\right]^2}{\iiint (S_Q + S_E + MTF^2 S_B) MTF^2 W^2_{\text{task}} df_x df_y df_z}. \tag{7}$$

The NPW model has been shown to describe human observers somewhat better than the PW model in cases of, for example, high-pass-filtered images, such as CT.[10,46]

II.C.3. 2D "slice" models including the eye filter and internal noise (PWEi and NPWEi)

The two models above can be extended to include an eye filter, E, and observer internal noise, N_i.[47,48] While the form of the eye-filter and internal noise models have been investigated for 2D image data, they have not been extended or validated with regard to 3D data (e.g., volumetric viewing, scrolling of slices) and fully 3D observer models are a subject of ongoing research within the perception community. The fully 3D forms for d'_{PW} and d'_{NPW} assume the observer is capable of perceiving the entirety of volumetric image information, which may or may not be the case for scrolling slices, etc. However, the fully 3D d' can be reduced to a form describing detectability in a single (axial, coronal, or sagittal) slice extracted from a 3D volume as described by Gang *et al.*[10] Since extraction of a slice in the spatial domain [e.g., in the axial (x,y) domain] corresponds to convolution with the Fourier profile of slice thickness in the orthogonal direction of the Fourier domain [e.g., in the f_z direction], the slice detectability index can be written with an extra integral (convolution) over the orthogonal domain. For example, assuming an infinitesimally thin slice, the axial slice detectability index is given by convolution over f_z (i.e., convolution with the Fourier transform of the delta function)

$$d'^2_{\text{slice},PW} = \iint \frac{\left(\int MTFW_{\text{Task}} df_z\right)^2}{\int (S_Q + S_E + MTF^2 S_B) df_z} df_x df_y, \tag{8}$$

$$d'^2_{\text{slice},NPW} = \frac{\left[\iint \left[\int MTFW_{\text{task}} df_z\right]^2 df_x df_y\right]^2}{\iint \left[\int (S_Q + S_E + MTF^2 S_B) df_z\right] \left[\int MTFW_{\text{task}} df_z\right]^2 df_x df_y}, \tag{9}$$

and similarly for the sagittal or coronal slice detectability with integrals over f_x or f_y, respectively. The detectability can now be written in a form including the 2D eye filter (E) and internal noise models (N_i). For the NPW model, for example, we have

$$d'^2_{\text{slice},NPWEi} = \frac{\left[\iint E^2 \left[\int MTFW_{\text{Task}} df_z\right]^2 df_x df_y\right]^2}{\iint \left(E^4 \left[\int (S_Q + S_E + MTF^2 S_B) df_z\right] \left[\int MTFW_{\text{Task}} df_z\right]^2 + N_i\right) df_x df_y}, \tag{10}$$

where the eye filter was taken from Barten[48]

$$E(f) = f \cdot e^{-cf}, \tag{11}$$

with $c = 2.2$ mm^{-1}, and the internal noise was scaled as a function of viewing distance (e.g., $D = 50$ cm) and the zero-frequency component of the white noise-equivalent of total NPS[49]

$$N_i = 0.001\left(\frac{D}{100}\right)^2 \left[NPS_{eq}(0)\right], \tag{12}$$

where NPS_{eq} is a white noise-power spectrum equivalent in total power to the total NPS (i.e., sum of S_Q, S_e, and S_B), and the scale factor 0.001 as in Gang *et al.*[10] The slice NPWEi model (i.e., the NPW model with eye filter and internal noise) has been shown to yield reasonable agreement with human observer response[10] for simple tasks over a broad range of imaging conditions.

II.C.4. Imaging tasks

The task function explicit within the detectability index conveys the spatial frequencies of interest in performing the task, most simply defined in terms of binary hypothesis testing—e.g., a detection task with hypotheses of "signal present" versus "signal absent" or a discrimination task of hypotheses "stimulus A" versus "stimulus B." Three imaging tasks were used in the work below, selected to represent a range in detail (i.e., low-, middle-, and high-frequency tasks) with general consideration of tissue contrasts relevant to extremity CT imaging (e.g., muscle, fat, and bone). The task function (W_{Task}) may be written as a product of a spatial-frequency term (F_{Task}) and a contrast term (C_{Task}) representing the signal difference of the stimulus against background

$$W_{\text{Task}}(f_x,f_y,f_z) = F_{\text{Task}}(f_x,f_y,f_z)C_{\text{Task}}, \tag{13}$$

$$C_{\text{Task}} = \left(\mu^{\text{eff}}_{\text{object}}(kVp) - \mu^{\text{eff}}_{\text{background}}(kVp)\right)\frac{1}{(1+SPR)}P_{\text{vol}}, \tag{14}$$

$$P_{\text{vol}} = \begin{cases} \left(\dfrac{a_{\text{task}}}{a_{\text{vox}}}\right)^3 & \text{if } a_{\text{task}} \le a_{\text{vox}} \\ 1 & \text{if } a_{\text{task}} > a_{\text{vox}} \end{cases}, \tag{15}$$

where the contrast term C_{Task} for CBCT images is fundamentally given by the difference in effective attenuation coefficients ($\mu^{\text{eff}}_{\text{object}} - \mu^{\text{eff}}_{\text{background}}$) each computed by integrating the linear attenuation coefficients $\mu(E)$ over the x-ray spectrum and including two factors of contrast degradation: x-ray scatter $[1/(1+SPR)]$ and partial volume (P_{vol}) effect. The spectrum was modeled using the Spektr toolkit[50] implementation of the TASMIP model.[51] The attenuation coefficients were taken from NIST[52] and ICRU Report 46.[53] With regard to an extremity CT scanner, stimuli were assumed to be at the center of an object approximating a knee—i.e., a 16 cm water cylinder including an inner cylinder of 8 cm diameter bone (viz., 7.6 cm of trabecular bone and 0.4 cm of cortical bone).

The contrast term in each task function carries a dependence on kVp, SPR, system magnification, and a possible partial volume effect. The kVp dependence is implicit in the effective attenuation coefficients. The SPR dependence is manifest in the factor $1/(1+SPR)$, which in turn depends on magnification as in Appendix B. The partial volume factor describes a loss in signal magnitude for stimuli smaller than the voxel size. An important caveat regarding partial volume effect must be noted: P_{vol} calculations assume that the 3D stimulus is located at the center of a single voxel, and that the signal is not spread across multiple voxels. This assumption violates the shift invariance assumption underlying cascaded systems analysis, invoking the "cyclostationary" condition suggested by Cunningham *et al.*[54] and recognizing that the effects of such shift invariance is typically small.[55] In this way, stimuli smaller than the voxel size (and assumed to reside entirely within a single voxel) present as a single enhanced voxel but with contrast reduced by the P_{vol} factor.

The frequency term was computed separately for three tasks emphasizing low, middle, and high frequencies. Task 1 was a low-contrast ("muscle stimulus vs water background") task emphasizing low spatial frequencies in the detection of a Gaussian blob on a uniform background

$$F_{\text{Task}}(f_x,f_y,f_z) = (\sqrt{2\pi}a_{\text{task}})^3 e^{-2\pi^2 a^2_{\text{task}}(f_x^2+f_y^2+f_z^2)}, \tag{16}$$

with the size of the stimulus, a_{task}, taken nominally as 1 mm. For task 1, $\mu^{\text{eff}}_{\text{object}}$ was computed from the attenuation coefficient of muscle, and $\mu^{\text{eff}}_{\text{background}}$ was computed from the attenuation coefficient of water.

Task 2 was a high-contrast ("bone vs fat") task emphasizing middle frequencies corresponding to detection of a submillimeter cube (3D rect) against a uniform background

$$F_{\text{Task}}(f_x,f_y,f_z) = \begin{cases} a^3_{\text{task}}.\text{sinc}(f_x a_{\text{task}})\text{sinc}(f_y a_{\text{task}})\text{sinc}(f_y a_{\text{task}}) \text{if } a_{\text{task}} > a_{\text{vox}} \\ a^3_{\text{vox}}.\text{sinc}(f_x a_{\text{vox}})\text{sinc}(f_y a_{\text{vox}})\text{sinc}(f_y a_{\text{vox}}) \text{if } a_{\text{task}} > a_{\text{vox}} \end{cases}, \tag{17}$$

where the first conditional corresponds to a stimulus greater than the voxel size, and the second implies that the stimulus is smaller than the voxel size and is evident as an enhancement of the entire voxel (but at reduced contrast due to partial volume). The nominal stimulus size was $a_{\text{task}} = 0.3$ mm, and the voxel size varied as (a_{pix}/M). For task 2, $\mu^{\text{eff}}_{\text{object}}$ was computed from the attenuation coefficient of bone, and $\mu^{\text{eff}}_{\text{background}}$ was computed from the attenuation coefficient of fat.

Task 3 was a low-contrast ("muscle vs water") task emphasizing high spatial frequencies corresponding to discrimination of a Gaussian against statistically defined "white" noise background

$$F_{\text{Task}}(f_x,f_y,f_z) = 1 - e^{-2\pi^2 a^2_{\text{task}}(f_x^2+f_y^2+f_z^2)}, \tag{18}$$

where the size of the stimulus was $a_{\text{task}} = 0.65$ mm. For task 3, $\mu^{\text{eff}}_{\text{object}}$ and $\mu^{\text{eff}}_{\text{background}}$ were the same as for task 1.

II.D. Optimization parameters

The optimization study was organized according to: (1) factors affecting spatial resolution; (2) factors affecting image noise and dose; and (3) multivariate combinations

thereof. The range and nominal values for the most pertinent parameters are summarized as follows.

II.D.1. Factors of spatial resolution

II.D.1.a. Magnification System magnification ($M = SDD/SAD$) varied from 1 to 2, keeping source-to-isocenter distance (SAD) fixed at 410 cm to fix the dose at isocenter irrespective of magnification. The nominal value was $M = 1.28$, corresponding to the value identified in preliminary work.[1,2] Variation in M affected a multitude of dependent parameters, including: focal spot and detector MTF (referenced to isocenter); fluence to the detector (inverse-square law); the "natural" voxel size $\left(a_{\text{vox}} = a_{\text{pix}}/M\right)$ and the SPR (via the scatter point source model in Appendix B).

II.D.1.b. Focal spot size X-ray focal spot size varied from 0 to 1.5 mm, with a nominal value of $a_{\text{spot}} = 0.5$ mm corresponding to a commercially available fixed-anode x-ray tube identified in preliminary work (Source-Ray XRS-125-7K-P, Source-Ray, Ronkonkoma, NY).

II.D.1.c. Pixel and voxel size Pixel aperture ranged from 0.05 to 0.40 mm, typical of the range in FPDs. The nominal value was $a_{\text{pix}} = 0.194$ mm (corresponding to the PaxScan 3030+, Varian Imaging Products, Palo Alto, CA). For real FPDs, the fill factor depends on the pixel size due to finite feature size constraints in the fabrication process: for a_{pix} approaching 0.05 mm, the fill factor could be less than 50%, while for a_{pix} approaching 0.40 mm, the fill factor is ~80%. For simplicity, we held the fill factor fixed at a nominal value of 0.68 in all cases. The voxel size was nominally calculated as the natural voxel size, $a_{\text{vox}} = a_{\text{pix}}/M$, giving nominal value of $a_{\text{vox}} = 0.15$ mm for $a_{\text{pix}} = 0.194$ mm and $M = 1.28$.

II.D.1.d. Reconstruction filter The apodization cosine filter parameter h_{win} (Appendix A) varied from 0.5 to 1.0, corresponding to the range from Hann to ramp (Ram-Lak), respectively.

II.D.2. Factors of image noise and dose

II.D.2.a. kVp Peak kilovoltage ranged from 60 to 140 kVp, with nominal value of 90 kVp identified in preliminary work.[1,2] Total filtration was fixed at 5.1 mm Al + 0.3 mm Cu. Filter thickness was based on the result in the companion paper (Ref. 2), which optimized 2D projection DQE as a function of kVp and added Cu filtration and determined a reasonable filter thickness of 0.3 mm, subject to x-ray tube operating power and heating constraints. Regarding the choice of material (e.g., Cu and Ag) sample calculations and previous experience suggest that filter material is not likely a parameter of first-order importance in single-energy CBCT. A body of previous work[50,56,57] shows optimal filters that tend toward thick, high-Z materials, essentially monochromatizing the beam. Such optimal filters would carry high heat load requirements beyond the limits of the current system, as addressed partially in the companion paper (Ref. 2). Therefore, for the current work on single-energy CBCT modeling, added beam filtration was held fixed. In future work regarding dual-energy (DE) CBCT, the addition of differential filters in the high- and low-energy beams is recognized as important parameter[58] that should be considered in system optimization.

Variation in kVp affected a multitude of dependent parameters, including patient dose, detector DQE (i.e., $\bar{g}_1$ $\bar{g}_2$, etc.), fluence at the detector, anatomical background power spectrum (via the κ term in S_B), and contrast [via attenuation coefficients $\mu(E)$]. The range of deliverable kVp (and mA at a particular focal spot size) for a particular tube and generator depends on the available power (kW). Since the analysis below was intended to be fairly general, we allowed kVp to vary freely without constraining the power to that of a particular system, but the kW required to deliver a given kVp and mA (dose) was identified.

II.D.2.b. Dose Absolute dose to the center of a 16 cm diameter water cylinder was varied from 0.1 to 10 mGy, with a nominal value of 5 mGy identified in preliminary work as appropriate to extremity CT.[1] Dose was related to the exposure (X) at the detector after attenuation by the object as

$$\text{Dose} = mX\left(\frac{SDD}{SAD}\right)^2 e^{\left(\mu^{\text{eff}}_{\text{water}}R\right)} f_{\text{water}} BSF, \tag{19}$$

where $(SDD/SAD)^2$ corrects for inverse-square law, m is the total number of projections, R is the radius of the water cylinder ($2R = 16$ cm), f_{water} is the f-factor for water ~0.9 cGy/mR (from Hubbell and Seltzer[52] averaged over the x-ray spectrum), and BSF ~4.5 is the backscatter factor experimentally determined on an x-ray bench (described in Sec. III A) for the proposed scanner geometry.

II.D.2.c. FPD electronics noise The FPD readout electronics noise, σ_{add}, ranged from 0 to 5000 e, with a nominal value of 2000 e corresponding to the PaxScan 3030+ FPD (Varian Imaging Products, Palo Alto, CA).

III. EXPERIMENTAL METHODS

III.A. Validation of measured and predicted noise-power spectrum

Such theoretical modeling has been widely employed in a variety of applications and was validated below for the nominal detector, system geometry, etc., associated with the extremity CBCT scanner. Previous work by Tward and Siewerdsen [5] demonstrated agreement of theoretical and measured 3D NPS in CBCT using the same underlying model, and similar validation is confirmed in the current work for a system differing significantly in the choice of system geometry, x-ray tube, and FPD. That is, the previous validation in the context of a system for CBCT-guided radiation therapy deserved re-examination in the context of a system for CBCT of MSK extremities. Given agreement between theory and measurement in the 3D NPS, Gang *et al.*[10] extended the analysis to demonstrate reasonable correspondence between task-based detectability index (which is ultimately derived from the NPS) and the performance of human observers for a broad range of experimental conditions and simple imaging tasks. Thus, the work of Tward *et al.* in combination with that of Gang *et al.* supports a rationale in which validation of the 3D NPS model combined with simple task-based analysis of detectability index should yield results providing reasonable correspondence with real

observer performance for simple imaging tasks. We therefore validated the cascaded systems model in terms of the 3D NPS but refrained from the (far more challenging) revalidation versus real observer performance. CBCT scans were acquired on an experimental imaging bench for tomosynthesis and CBCT having a fixed tube-detector and rotating object configuration on a system of translation and rotation stages. The operation of the bench is similar to that detailed in Tward and Siewerdsen .[5] The system consisted of a PaxScan 4030CB panel (Varian Imaging Products, Palo Alto, CA) with 2048×1536 pixels at 0.194 mm pitch and a 250 mg/cm^2 CsI:Tl converter. Model parameters for the 4030CB were equivalent to the 3030+ FPD, differing only in field of view (FOV). The x-ray tube was a DU694 in EA10 housing (Dunlee, Aurora, IL) with a 14° anode angle, focal spot size of 0.4 or 0.8 mm and peak tube voltage of 150 kVp. Motion control was provided by Compumotor 6k8 (Parker Hannifin, Cleveland, OH) and system geometry (magnification) was set to closely match the proposed scanner setup with focal spot size fixed at 0.5 mm.

CBCT images were acquired for measurement of the 3D NPS as a function of: kVp (70, 90, and 110 kVp), tube current (0.25, 0.5, and 0.8 mAs), and reconstruction filter window (Hann or ramp). Acquisition settings for a "nominal" CBCT scan were fixed at 90 kVp; 0.3 mm Cu + 4 mm Al added filtration; 0.25 mAs per projection; 360 projections across 360° circular orbit with uniform Parker weighting; $SAD = 434$ mm; $SDD = 555$ mm; with other parameters described in Table I. CBCT images of air (no phantom) were acquired with 3 mm Cu filtration added to the beam to simulate x-ray attenuation similar to a 16 cm water cylinder. The absence of an object in the FOV avoided complications arising from nonuniform x-ray scatter, beam hardening, etc. Dark flood corrections were performed using 50 dark frames and 50 flood-field images acquired immediately before image acquisition with no object in the field of view.

The projections were reconstructed using the FDK algorithm.[59] For each set of acquisition settings, two scans were repeated and the reconstructed volumes were subtracted. This operation removed spatial correlations (e.g., shading artifacts) between the two scans, and the resulting 3D difference image presented stochastic noise components (quantum and electronics noise), contributing independently between the scans. The NPS was measured from the ensemble average of the squared discrete Fourier transform (*DFT*) of volumes of interest (VOIs) in the difference image as follows[5]:

$$S_{\text{measured}}(f_x,f_y,f_z) = \frac{1}{2}\frac{\Delta_x}{N_x}\frac{\Delta_y}{N_y}\frac{\Delta_z}{N_z}\left\langle |DFT\{VOI_{\text{sub}}(x,y,z)\}|^2 \right\rangle, \quad (20)$$

where Δ_i denotes voxel size (mm) in the i direction, N_i (= 71) is the number of voxels in each VOI in the i direction. $VOI_{\text{sub}}(x, y, z)$ is the mean-subtracted volume of interest in the difference image. The VOI in the difference image was found to have a very small, residual mean signal on the order of 10^{-7} cm^{-1} (units of attenuation coefficient). While this was likely an insignificant departure from a true zero-mean, we further performed mean subtraction on the VOI to obtain a signal as close to zero-mean as possible. VOIs (each $71 \times 71 \times 71$ voxels) in the difference image were selected about two concentric circles (36 VOIs on each circle) at radii of 60 voxels and 80 voxels from the center of the volume. This avoided placing VOIs near the center (ring artifacts) or the edge (region outside circle of reconstruction) of the image volume. The factor of two in the denominator corrects for the subtraction of two statistically independent scans.

For display purposes, a radial average of the measured NPS about the f_z axis was computed. The results show a 1D profile of the NPS taken slightly (one voxel) above the axial plane to avoid on-axis artifacts. Error bars were twice the standard deviation of samples in each frequency bin and normalized by the square root of the number of samples in each bin. Radial averaging reduced the statistical error in plots and condensed a large amount of 3D information to a form that was easily displayed. Since the NPS is not strictly radially symmetric, displaying the plots in this manner slightly distorts the underlying frequency dependence, but it still provides a fair comparison of theory with measurement (both slightly distorted by the radial average). The 3D NPS is reported with units of $(\mu^2)(mm^3)$, where μ refers to linear attenuation coefficient and carries units of mm^{-1}.

III.B. Qualitative verification in phantom/cadaver studies

CBCT scans of a cadaver hand and knee were acquired on the imaging bench to provide a qualitative assessment of theoretical calculations. Three cases were investigated for qualitative comparison to the predicted dependence on: kVp (60 and 105 kVp, each with 5.1 mm Al + 0.3 mm Cu filtration, and mAs per projection adjusted to give 10 mGy for each scan (confirmed with a Farmer ionization in a 16 cm CTDI phantom and a Radcal dosimeter); pixel aperture [a_{pix} in full-resolution mode (0.194 mm) and half-resolution mode (0.388 mm)], and reconstruction filter window [$h_{\text{win}} = 0.5$ (Hann) and 1.0 (Ram-Lak)]. Nominal geometry and other settings were as described in Sec. III A.

IV. RESULTS

Results in the following sections include: (1) comparison of the theoretical model for 3D NPS to measurement under conditions corresponding to the proposed musculoskeletal scanner; (2) calculations of detectability index for various observer models; (3) analysis of the impact of background clutter; and (4) multivariate analysis of detectability as a function of factors affecting spatial resolution, noise, and dose.

IV.A. Comparison of theory and measurement (NPS)

The theoretical and measured NPS are shown in Fig. 2. Figure 2(a) shows a strong decrease in NPS for higher kVp (at fixed mAs and, therefore, variable dose), attributed primarily to increased fluence at the detector. Figure 2(b) demonstrates the expected inverse relationship between NPS and mAs (at fixed kVp). Figure 2(c) shows the increase in NPS

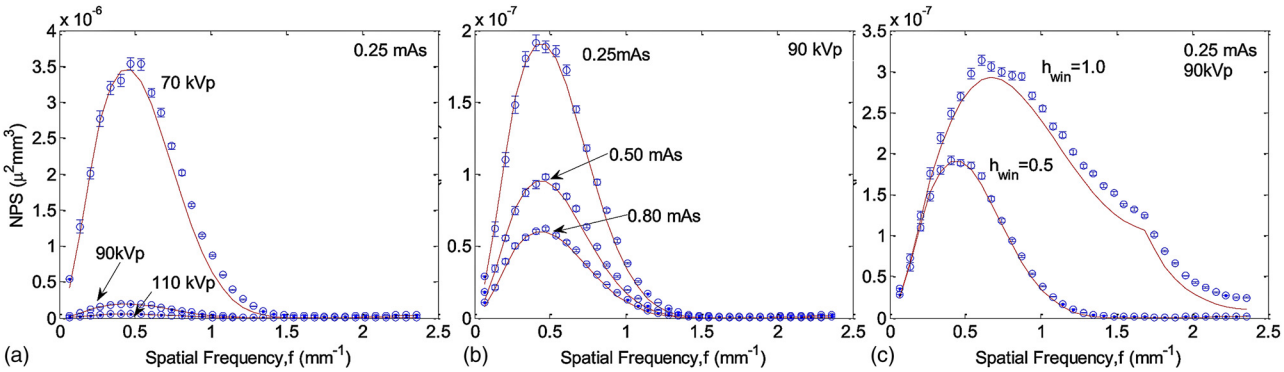

FIG. 2. Comparison of experimental and theoretical 3D NPS for various (a) kVp (70, 90, and 110 kVp at a fixed tube current of 0.25 mAs /projection), (b) mAs (0.25, 0.50, and 0.80 mAs per projection at a fixed tube voltage of 90 kVp), and (c) reconstruction filter ($h_{\text{win}} = 0.5$ Hann and $h_{\text{win}} = 1$ Ram-Lak). Each plot shows a radial average of the central axial slice of the fully 3D quantum NPS ($S_Q^{\text{tot}} + S_E^{\text{tot}}$) represented by Eqs. (C3) and (C5).

for sharper apodization filters (Hann versus Ram-Lak). Excellent agreement is demonstrated between theory and measurement overall, with a slight systematic underestimate in the case of the sharp (Ram-Lak) filter due to a numerical approximation of the 3D sampling process (stage 14, computed by summation of 26 nearest-neighbor replicants of the 3D NPS at multiples of the sampling frequency rather than a full 3D convolution). The underestimate is more evident for the Ram-Lak filter, since the sharper filter suffers a greater 3D noise aliasing effect. The results are consistent with the level of agreement demonstrated in previous work[5,6] and establishes a basic validation of the quantum noise model under conditions appropriate to the musculoskeletal scanner (i.e., system geometry, detector selection, etc.).

IV.B. Observer models

Figure 3 shows calculations of the detectability index for five observer models as described by Eqs. (6)–(10), specifically 3D PW, 3D NPW, 2D PW axial, 2D NPW axial, and 2D NPWEi axial. In each case, calculations were performed for task 1 (low-frequency detection task) and task 3 (high-frequency discrimination task) as a function of geometric magnification and detector pixel size. The low-frequency task (top row) consistently suggests higher performance at lower magnification and larger pixel size—consistent with intuition for detection of a low-frequency Gaussian. The high-frequency task demonstrates an optimum in the range of $M \sim 1.2$–1.3 and $a_{\text{pix}} \sim 0.1$–0.2 mm. Although each of the five observer models suggests the same overall trend, there are appreciable differences. First, the magnitude of d' is seen to decrease as expected from the ideal 3D PW observer to the various NPW, 2D slice, and eye-filter models. Moreover, the optima are a bit steeper for the latter models—e.g., the dependence on M for the low-frequency task—suggesting increased susceptibility to quantum noise.

For brevity, we limit the results displayed below to one observer model that conveyed the overall trends. Recent work[10] examined the degree of correspondence between d' and real observers and found the 2D NPWEi slice model to give reasonable agreement for a variety of simple tasks. The slice models, however, require selection of a given (axial, sagittal, or coronal) slice in Eqs. (8)–(10) (i.e., the integral over f_z, f_y, or f_x, respectively), and we did not want to limit the results to a

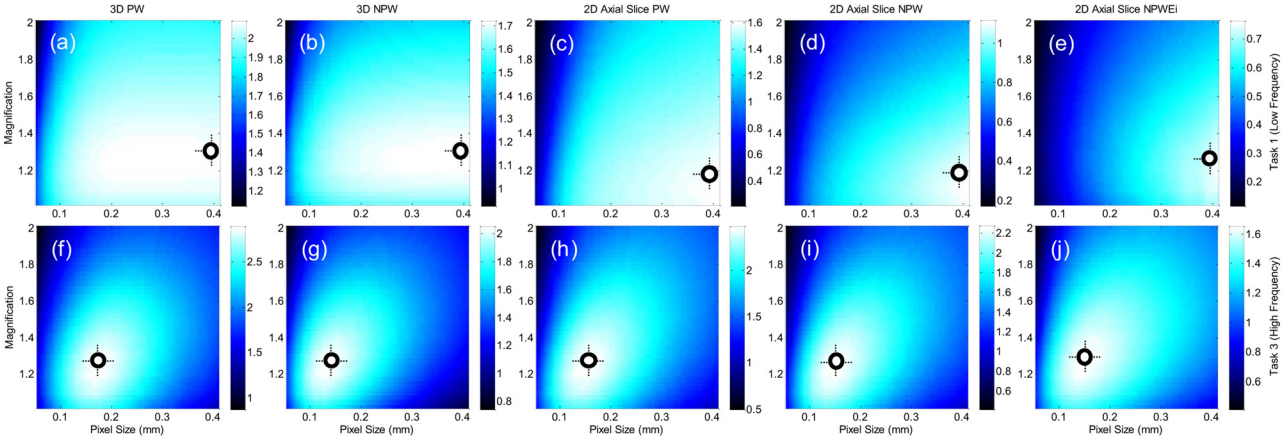

FIG. 3. Detectability index computed for the five observer models in Eqs. (6)–(10) as a function of magnification and detector pixel size (a_{pix}). (Top row) $d'(M, a_{\text{pix}})$ for task 1 (low-frequency muscle vs water detection task). (Bottom row) $d'(M, a_{\text{pix}})$ for task 3 (high frequency muscle vs water discrimination task). The low-frequency task exhibits lower d' overall due to background anatomical noise and is clutter-limited to a greater extent. The high-frequency task emphasizes detail and is less susceptible to clutter. Both tasks suggest highest performance at $M \sim 1.2$–1.3. The low-frequency task favors large pixel size, as expected, whereas the high-frequency task demonstrates an optimum $a_{\text{pix}} \sim 0.1$–0.2 mm.

preferred plane of visualization. We, therefore, selected the 3D NPW model as the nominal choice in results summarized below, since: (a) it captured the fully 3D characteristics of the noise (as opposed to selecting a given slice), (b) it incorporated the nonprewhitening aspects that appear to be important in correspondence with real observers, and (c) it was consistent with the overall trends suggested by all models in Fig. 3 (i.e., it did not suggest optima that were significantly different from the others). The choice of 3D NPW in results summarized below is purely for brevity, and the analysis was performed (but not shown) for all five. While the d' for the 3D NPW model is not expected to be absolutely predictive of real observer response (e.g., scrolling slices) in its magnitude, the trends and optima were consistent with those suggested by the other models and provided a concise, representative summary of results.

IV.C. Background anatomical clutter

Figure 4 illustrates the impact of anatomical clutter on d'. Figure 4(a) plots d' as a function of dose to isocenter for the three tasks with background clutter included (solid lines, $\kappa = 3.3 \times 10^{-10}$ μ^2 mm^3, $\beta = 3$, corresponding to muscle-water clutter, as derived in Appendix D) or excluded (dashed lines, $\kappa = 0$). The middle- and high-frequency tasks (task 2 and task 3) are seen to improve steeply with dose and are, therefore, governed primarily by quantum noise. By comparison, the low-frequency task (task 1) departs from this trend above ~2 mGy, beyond which d' increases slowly with dose, demonstrating that the task is largely "clutter-limited." The result is consistent with the notion that background clutter presents a major impediment to low-frequency tasks (e.g., soft-tissue mass detection) and outweighs other noise sources (quantum and electronics). The medium (task 2) and high (task 3) frequency tasks place emphasis outside the low frequency clutter region and are impervious to background clutter. (The solid and dashed lines in Fig. 4 are nearly identical for tasks 2 and 3.).

To illustrate the difference in dependence on anatomical clutter between low- and high-frequency tasks, d' was computed as a function of clutter magnitude (κ) and correlation (β) in Figs. 4(b) and 4(c). β varied from 1 to 5, whereas κ ranged in magnitude from 10^{-7}(corresponding to measurements of clutter in a phantom consisting of acrylic-in-air[30]) to 10^{-10} (calculated for clutter consisting of muscle-in-water). The low-frequency task [Fig. 4(b)] exhibits a strong dependence on both κ and β, whereas the high-frequency task [Fig. 4(c)] is relatively insensitive to clutter (governed instead by quantum noise).

IV.D. Factors of spatial resolution

Among the parameters governing spatial resolution are the system geometry (M), focal spot size (a_{spot}), pixel size (a_{pix}), voxel size (a_{vox}), and reconstruction filter window (h_{win}). As shown above [Fig. 3(g)], d' (M, a_{pix}) exhibited an optimum at $M \sim 1.3$ and $a_{pix} \sim 0.1$–0.2 mm for the middle- and high-frequency tasks. The optimal magnification arises from the interplay between factors of focal spot size, detector exposure, pixel (voxel) size, and SPR—the first two favoring lower magnification, and the latter two favoring higher magnification. For the low-frequency task 1, a similar optimum $M \sim 1.3$ was found, but there was no optimum in a_{pix} (favoring monotonically larger pixel size) over the range considered. With respect to magnification, the fact that a similar optimum was found for widely varying tasks is convenient from the perspective of system design—arguing that $M \sim 1.3$ is optimal for both the geometric magnification required in high-frequency tasks as well as the SPR reduction required for low-frequency tasks. The lack of a single optimum in a_{pix} suggests a separate approach in image acquisition/reconstruction: for high-frequency tasks, projection data should be acquired near the optimum $a_{pix} \sim 0.1$–0.2 mm; however, for low-frequency tasks, the same data should be binned to $a_{pix} \sim 0.4$ mm and reconstructed separately.

Figure 5 illustrates the corresponding trends in d' computed as a function of reconstruction filter (h_{win}) and focal spot size (a_{spot}). Calculations of d' (M, h_{win}) in Figs. 5(a)–5(c) show a consistent optimum at $M \sim 1.3$ for each task. The low-frequency task shows a weak dependence on reconstruction filter but favors the smoother filter ($h_{win} \sim 0.5$), whereas the middle- and high-frequency tasks show a steeper dependence and favor the sharper filter ($h_{win} \sim 0.9$–1.0). As shown in Figs. 5(d)–5(f), calculation of d' (a_{pix}, h_{win}) for the low-frequency

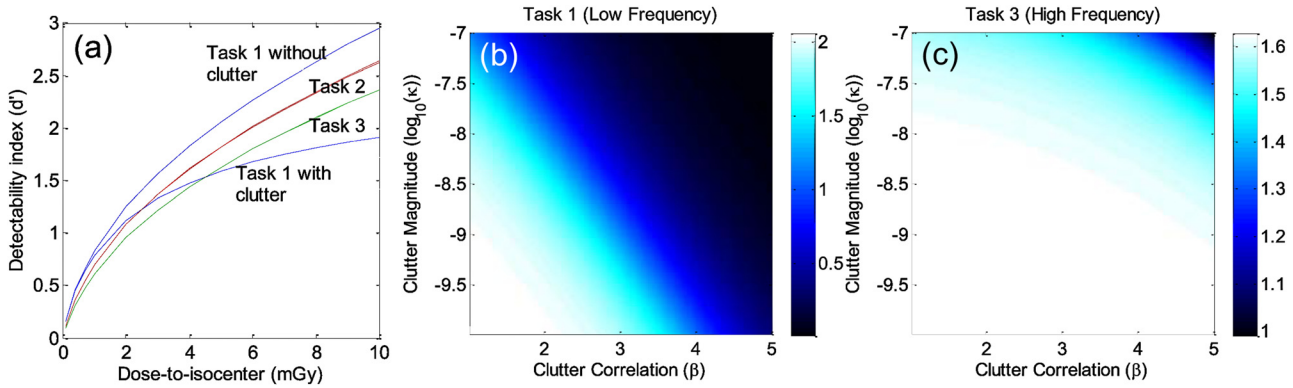

FIG. 4. Effect of background anatomical clutter on detectability. (a) d' computed as a function of dose for the three tasks with (solid lines) and without (dashed lines) background clutter in Eq. (7). For task 1 with clutter, the departure from quantum-limited performance (i.e., relative insensitivity to dose above ~2 mGy) shows that the low-frequency task is clutter-limited. Tasks 2 and 3 are insensitive to clutter (solid and dashed lines overlapping). (b) $d'(\kappa, \beta)$ for low-frequency task 1 exhibits a strong dependence on clutter, whereas (c) $d'(\kappa, \beta)$ for high-frequency task 3 is relatively insensitive to clutter.

task again favors larger pixels and smoother filters, whereas the middle- and high-frequency tasks show an optimum $a_{pix} \sim 0.1$–0.2 mm and favor sharp filters. The trends are consistent with intuition and suggest a separate reconstruction strategy for low versus middle/high-frequency tasks: binning and smooth filters for the former; full-resolution readout and sharp filters for the latter. While intuitive, the results confirm expected trends and quantify the trade-offs in detectability incurred by departures from the optima.

Finally, Figs. 5(g)–5(i) plot d' as a function of a_{pix} and a_{spot}, illustrating key factors of equipment selection—i.e., the FPD and x-ray tube. The familiar trends in a_{pix} are evident at all values of a_{spot} considered (i.e., $a_{pix} \rightarrow 0.4$ mm for the low-frequency task, while $a_{pix} \sim 0.1$–0.2 mm optimizes the middle- and high-frequency tasks). For the low-frequency task, there is a very weak dependence on focal spot blur. For the middle- and high-frequency tasks, however, there is a fairly steep, monotonic trend toward finer focal spot. However, at optimal a_{pix}, one may argue that decrease in a_{spot} below $\sim$0.4 mm provides diminishing return—an important consideration with respect to tube heat loading. At least as interesting, the plots allow a quantitative guide to system design questions—e.g., the conditions under which reducing pixel size or spot size in isolation improves performance (or has no effect at all) and the extent to which the best choice of these parameters depends strongly on one another and the imaging task.

IV.E. Factors of image noise and dose

Among the factors governing image noise (aside from the spatial resolution factors above) are kVp, dose, and additive electronic noise (σ_{add}). In the results below, d' was analyzed as a function of each, recognizing that the contrast factor (C_{Task}) in the task function tends to heavily weight the results toward lower kVp. An additional figure of merit was, therefore, computed—the integral of the NEQ [Eq. (5)] over the 3D Nyquist region—equivalent to a delta-function detection task but without contrast dependence and conveying the trends associated solely with quantum noise, electronic noise, and anatomical clutter (but not the contrast of the stimulus).

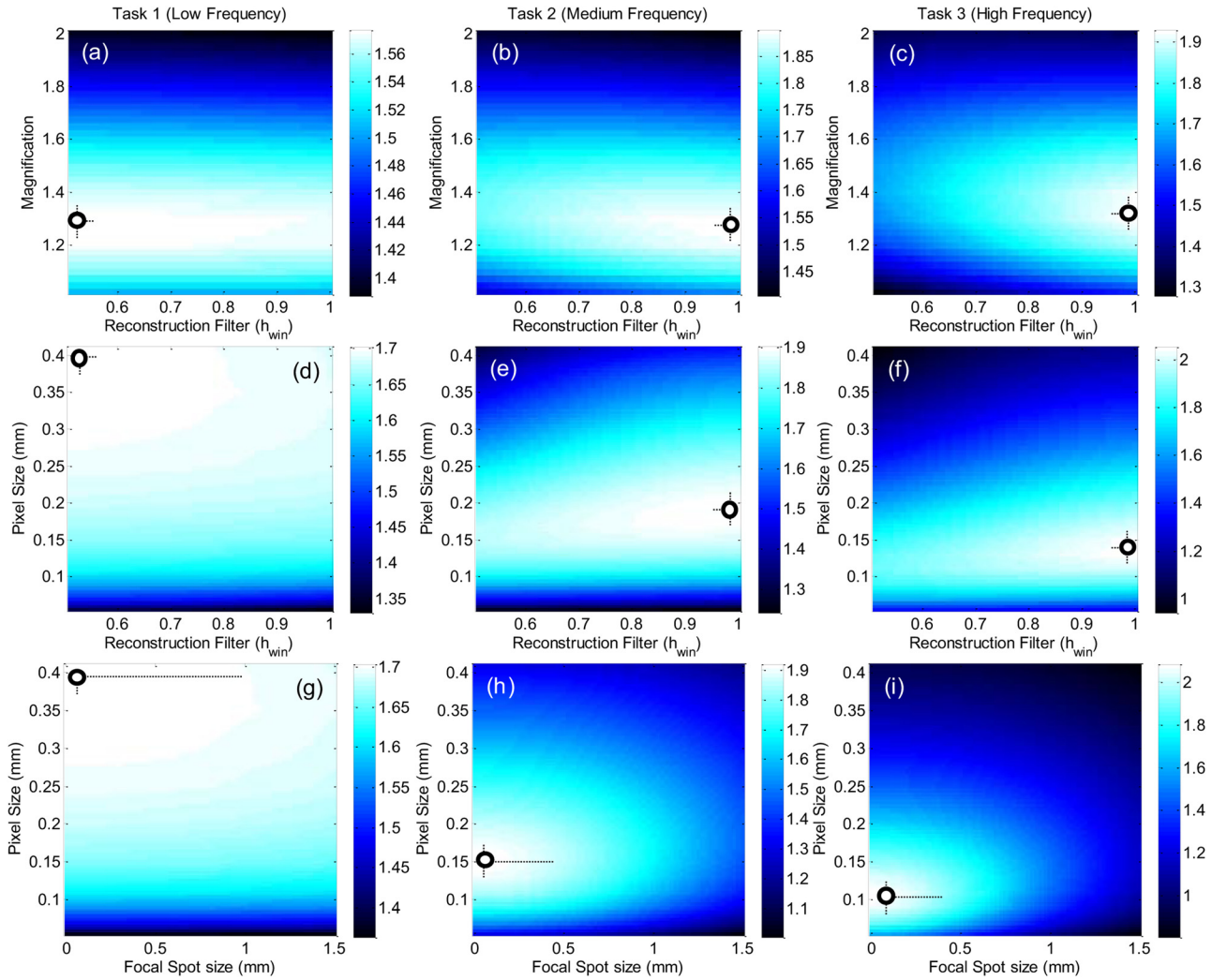

FIG. 5. Detectability computed as a function of factors intrinsic to spatial resolution in CBCT for three imaging tasks. (a–c) $d'(M, h_{win})$ demonstrates a consistent optimum at $M \sim 1.3$, a slight preference for smoother filters for the low-frequency task, and steeper dependence toward sharper filters for high-frequency tasks. (d–f) $d'(a_{pix}, h_{win})$ reflects the task-dependent effects of pixel size as in Fig. 3 and a similar dependence on reconstruction filter. (g–i) $d'(a_{pix}, a_{spot})$ shows monotonic improvement for smaller a_{spot}, with arguably diminishing returns below $a_{spot} \sim 0.4$ mm.

Figure 6 shows d' computed as a function of kVp, dose, and σ_{add}. While we do not expect an optimum in d'(dose, σ_{add}), levels of diminishing return become evident—e.g., the dose beyond which d' becomes entirely clutter-limited and does not increase appreciably with dose. Nominal operating points are, therefore, highlighted in each: dose ~5 mGy and $\sigma_{\text{add}} \sim 2000$ e, as identified in previous analysis.[1,2] The trends suggest kVp in the range of 60–80 kVp for all three tasks, recognizing that d' is strongly influenced by the C_{Task} term (which improves monotonically with lower kVp). The isopower contour lines in Figs. 6(a)–6(d) illustrate the need for careful selection of operating kVp: for a given dose, d' analysis suggests lower kVp but at the cost of higher kW. The dependence of the underlying noise on kVp (independent of the stimulus contrast) is illustrated in the integral of NEQ in Figs. 6(d) and 6(h), which suggests a fairly weak (slowly varying) optimum around ~90 kVp (at 5 mGy). The NEQ analysis highlights the important trade-off between quantum noise (S_Q) and anatomical clutter (S_B) with kVp: quantum noise alone tends to favor low kVp (higher quantum detection efficiency, $\bar{g}_1$), while higher kVp reduces κ and thus the clutter power spectrum. The same observation was made by Richard *et al.*[36] in the context of chest radiography, providing an analytical basis for weighing the trade-offs between quantum noise and clutter "contrast."

The dependence on kVp and additive noise (σ_{add}) is illustrated in Figs. 6(e)–6(h). Tasks 1 and 3 indicate an optimum ~70–80 kVp at the nominal electronic noise level ~2000 e, while task 2 (bone-fat contrast) suggests ~60 kVp, owing to steep dependence of the C_{Task} term on kVp. The integral of the 3D NEQ shows an optimum in the region of ~90 kVp.

IV.F. Multivariate optimization

Finally, the parameters governing spatial resolution and noise were considered together in Fig. 7. Calculations of $d'(M, \text{kVp})$ in the first row reflect the familiar optimum in the region $M \sim 1.3$. For the high-frequency task at $M = 1.3$, an optimal kVp is evident at ~70 kVp, while the integral of the 3D NEQ shows optima at $M \sim 1.2$–1.3 and ~90 kVp. Calculations of $d'(a_{\text{pix}}, \text{kVp})$ in the second row confirm the basic dependence on a_{pix} irrespective of kVp (i.e., $a_{\text{pix}} \to 0.4$ mm for the low-frequency task, compared to $a_{\text{pix}} \sim 0.1$–0.2 for the middle and high-frequency tasks). Calculations of $d'(a_{\text{pix}}, a_{\text{add}})$ in the third row demonstrate an interesting trade-off between pixel size and additive noise: for fixed electronic noise, reduction in pixel size increases quantum noise (i.e., electronics noise contribution grows by $1/a_{\text{pix}}^2$, Appendix C); hence, below a certain pixel size ($a_{\text{pix}} \sim 0.1$ mm), the increase in quantum NPS outweighs the benefits of improved spatial resolution, and an "optimal" pixel size is suggested. The low-frequency task 1 is relatively insensitive to σ_{add} and demonstrates diminishing returns [Fig. 7(i)] below $\sigma_{\text{add}} \sim 2000$ e at pixel size $a_{\text{pix}} \sim 0.4$ mm—i.e., low-frequency detection tasks are dominated by clutter and not electronic noise. Finally, we see that the results provide a guide to another important system design question with respect to a_{pix} and σ_{add} —namely, whether a reduction in pixel size alone will improve detectability, or whether a concomitant reduction in electronic noise is required. For example, considering the middle-frequency task 2 at $\sigma_{\text{add}} = 2000$ e, one might wonder if a reduction in pixel size below 0.15 mm would improve performance; the answer [Fig. 7(j)] is that an improvement would only be gained if the electronic noise were also reduced.

V. DISCUSSION AND CONCLUSIONS

The cascaded systems model of 3D quantum noise in CBCT was extended to include system parameters (magnification, focal spot blur, and x-ray scatter) and background anatomical clutter. Of these three general groups of factors affecting NEQ (i.e., quantum, system, and generalized factors as in Fig. 1), previous work has tended to treat one or two—e.g., analysis of CBCT quantum noise by Tward and Siewerdsen,[5] inclusion of scatter by Siewerdsen and Jaffray,[16] Jain *et al.*,[17] and

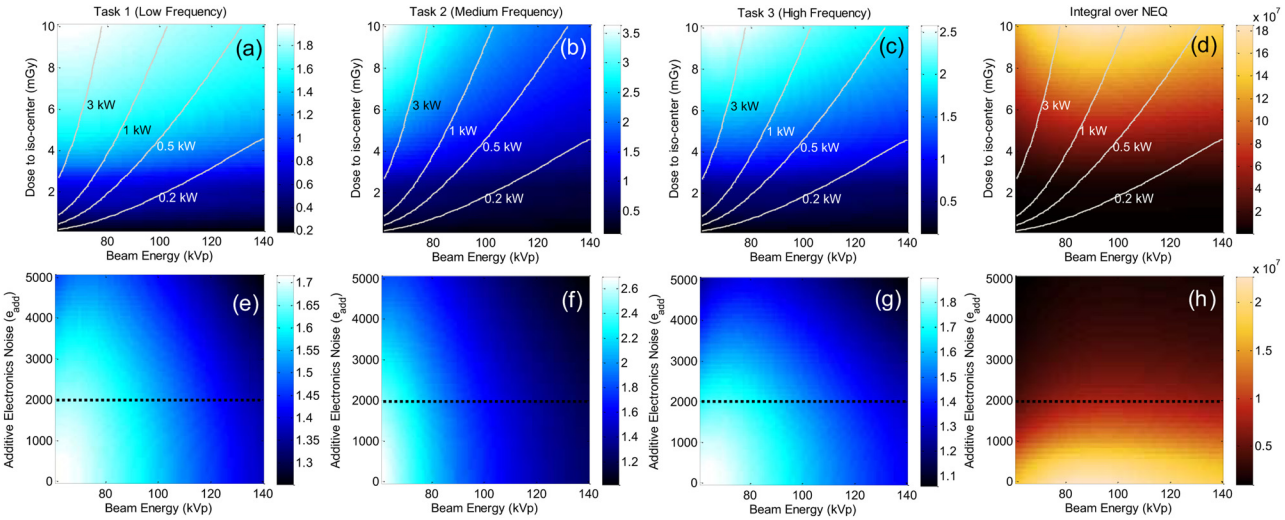

FIG. 6. Detectability as a function of factors contributing to image noise and dose for the three imaging tasks. The right column plots the integral over the 3D NEQ. The superimposed lines identify the generator power required for a given kVp and dose. While there is clearly no "optimum" in dose or σ_{add}, reasonable operating points of $D = 5$ mGy and $\sigma_{\text{add}} = 2000$ e are identified, and regions of diminishing return are evident.

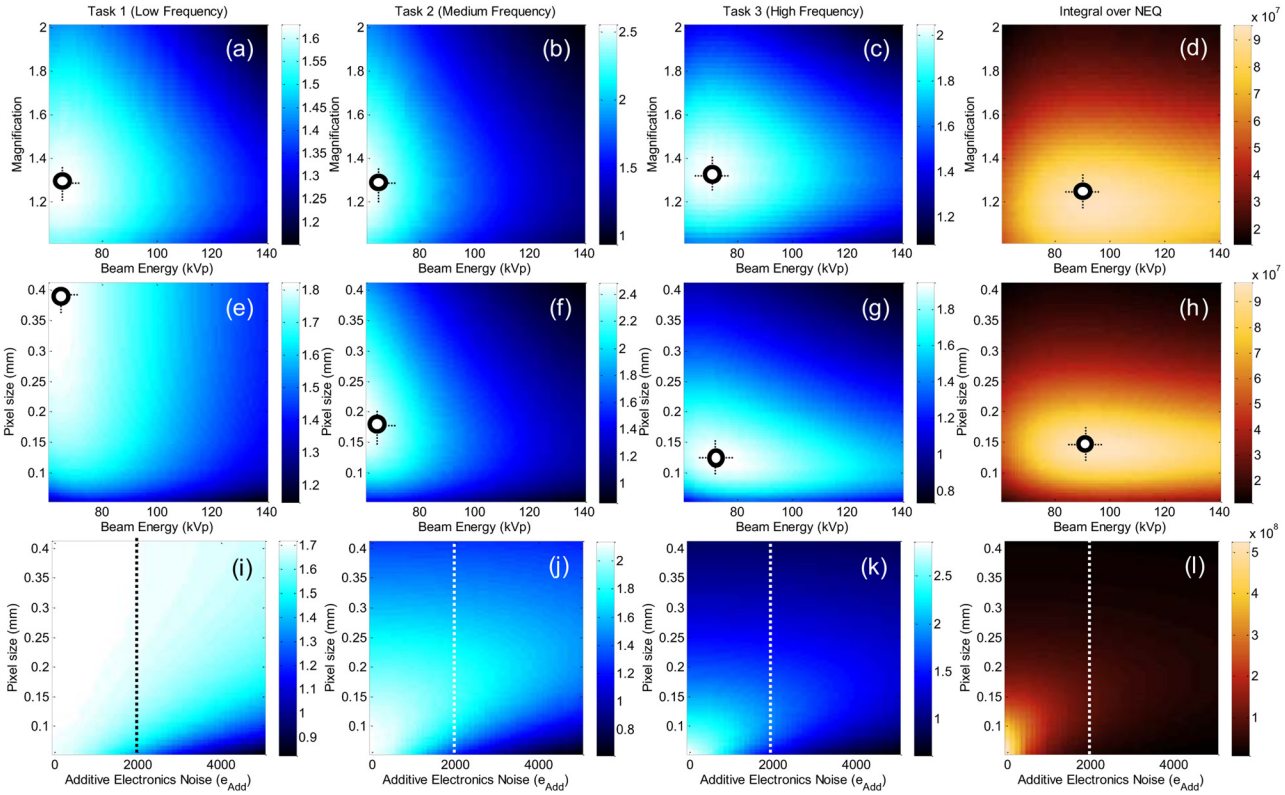

FIG. 7. Detectability index and integral of the 3D NEQ computed for parameters governing both spatial resolution and noise. (a–d) Dependence on magnification and kVp. (e–h) Dependence on pixel size and kVp. (i–l) Dependence on pixel size and electronic noise (σ_{add}).

Kyprianou *et al.*,[18] and incorporation of anatomical clutter as in Gang *et al.*[30] The model presented above combines all three to examine trade-offs and interdependencies among factors of system design, image acquisition, and 3D reconstruction and applies the framework specifically to a CBCT system under development for MSK extremity imaging.

A fairly broad multivariate parameter space was investigated systematically in terms of factors affecting: (1) spatial resolution (magnification, focal spot blur, reconstruction filter, pixel size, and voxel size); (2) image noise (kVp, dose, x-ray scatter, and electronics noise); and (3) each of these factors in combination. Among the specific findings for the MSK extremity scanner are: (1) A fairly robust optimal geometry was identified at $M = 1.3$, weighing trade-offs among spatial resolution (i.e., pixel size and focal spot size) and x-ray scatter. (2) Intuitive trends were quantified regarding acquisition/reconstruction techniques for low- and middle/high-frequency tasks—namely, 1×1 pixel readout (no binning; $a_{pix} \sim 0.15$ mm) followed by 1×1 or 2×2 binning in reconstruction depending on the imaging task (i.e., 2×2 pixel binning and a smooth filter for low-frequency, soft-tissue tasks, and a sharp filter with no binning for high-frequency, bone detail tasks). (3) Detectability index favored lower kVp, driven primarily by the contrast term in the task function, whereas a contrast-independent figure of merit (integral over the NEQ) suggested ~90 kVp as optimal and may better reflect the true performance limits associated with quantum noise for an intrinsically digital image (where display contrast may always be maximized via window/level). The 90 kVp setting is also better suited in terms of generator power requirement (<1 kW). (4) The generator power and focal spot size associated with these results are within practical considerations—e.g., a fixed-anode x-ray source identified in preliminary work (SourceRay XRS-125-7K-P, 0.875 kW, 0.5 mm spot). (5) The analysis guides selection of a nominal dose (~5 mGy) above which performance is primarily limited by anatomical clutter and below which is limited by quantum noise—suggesting an operating point that is clinically reasonable (perhaps advantageous) in comparison to conventional CT,[60] although direct comparison between absolute central dose (as in this paper) and CTDI (as in Biswas *et al.*[60]) can be problematic.

Having addressed such a broad parameter space from purely theoretical, quantitative terms, it is worthwhile to briefly consider the results in a complementary, qualitative sense. While a rigorous perception study is beyond the scope of the current work, an image can speak volumes, and it is useful to see the extent to which the calculations agree with the visual appearance of CBCT images. Figure 8 presents CBCT images of a cadaveric knee (a and b) and hand (c–f) acquired on the imaging bench under the nominal conditions summarized in Sec. III A. Figures 8(a) and 8(b) show a central axial slice acquired at 60 and 105 kVp. Consistent with the d' calculations (Figs. 6–7), low-frequency soft-tissue visibility (e.g., muscle-fat differentiation) is improved for the 60 kVp scan by virtue of improved soft-tissue contrast, and the zoomed insets similarly illustrate the improvement for a high-contrast detail (calcification). Consistent with the disparity between d' (which is maximized at 60 kVp) and integral of the NEQ (which is maximized at 90 kVp), the images also reflect a slight increase in quantum noise for the lower kVp, and it will

fall to future work to assess via controlled reader studies, which image carries superior diagnostic performance.

Figures 8(c) and 8(d) qualitatively illustrate the effect of pixel aperture (binning) on visualization of soft-tissue and bone detail. The 1×1 binning image [Fig. 8(c) with zoom inset] shows a fine level of detail in trabecular structure but suffers by comparison to the 2×2 binning image [Fig. 8(d)] in terms of soft-tissue visibility. Conversely, the latter suffers a significant loss of detail. While intuitive, these observations are consistent with the quantitative results in Figs. 5 and 7 that suggested these task-specific acquisition/reconstruction techniques. Finally, Figs. 8(e) and 8(f) show the effect of the reconstruction filter on soft-tissue and detail, where the Hann filter is seen to improve soft-tissue visibility, while the Ram-Lak filter is suited for bony details. The basic result is consistent with that of Fig. 5 not only in the obvious conclusion (i.e., smoother filters for soft-tissue visualization) but also in the more subtle observation that the effect of the reconstruction filter on task performance is not nearly as strong as the effect of pixel binning [Figs. 8(c) and 8(d)]. As points of discussion, these images lend qualitative interpretation to the strength (i.e., the "steepness") of the trends and optima identified in Figs. 3–7 and qualitative support beyond the purely quantitative results. They also give some confidence that the calculations will bear correspondence with real observer performance.

A variety of limitations in the method and results are acknowledged. Fundamental to cascaded systems analysis, NPS, and NEQ are the assumptions of stationarity and shift invariance, which are known not to hold strictly for 3D filtered backprojection. As in previous work, the analysis should be interpreted as a "local" description of noise characteristics within a region (e.g., a region about the center of reconstruction) for which image statistics are locally invariant. Nonidealities in detector response—e.g., image lag—were ignored, although such effects can be expected to affect the noise in CBCT. While motion blur (i.e., motion during x-ray exposure) was neglected, the proposed scanner employs pulsed exposures, which largely eliminates motion blur over the duration of each pulse (<30 ms).[2] As a simplifying assumption, we held the detector pixel fill factor fixed at 0.68 (corresponding to the fill factor on the FPD employed on the prototype) irrespective of the actual pixel size, which may not be a fair comparison considering the fact that smaller pixels tend to have smaller fill factors. Although the model accommodates MTF degradation due to scatter, the results involved the simplifying assumption that scatter only manifests as contrast degradation ($1/1+SPR$) and a loss of primary quanta (an effective increase in noise). Finally, the results do not constitute a full multivariate analysis or true optimization within the rich multidimensional parameter space, presenting instead two-dimensional slices through the multivariate space of kVp, M, a_{pix}, h_{win}, etc., holding other parameters fixed at nominal values identified based on repeated calculations (not shown) and reasonable operating points. A multivariate optimization was not performed for the following reasons: (1) The vast number of possible permutations over the CBCT parameter space creates a computationally intense problem, the results of which can be

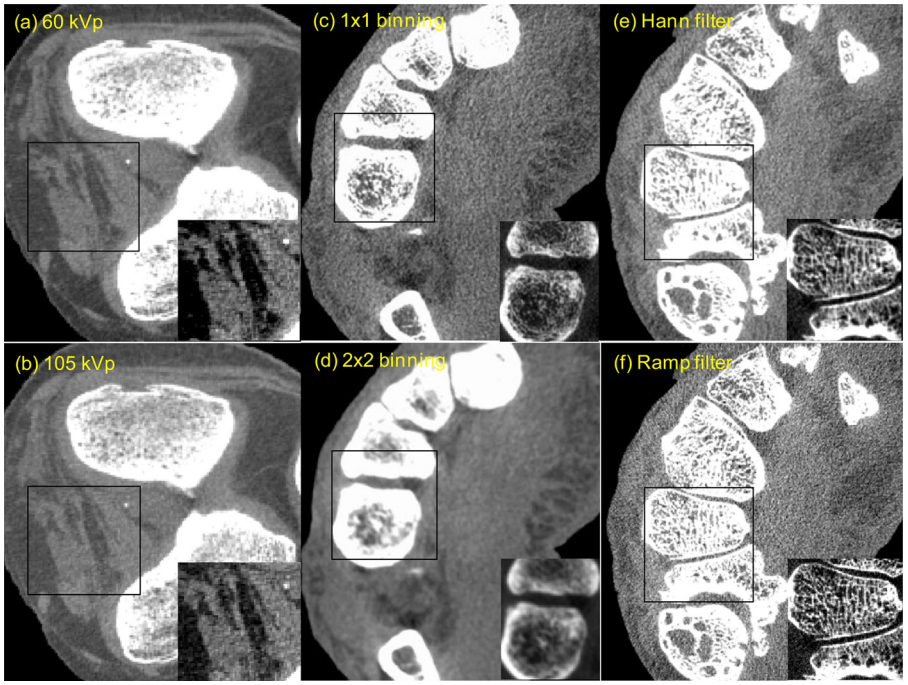

FIG. 8. CBCT images of (a and b) a knee and (c–f) a hand qualitatively illustrating some of the trends in d' theorized in Figs. 3–7. To facilitate comparison, each image pair is displayed at equivalent grey-scale window width, with level equal to the average voxel value for muscle. Each image also contains a magnified inset at separate window/level to illustrate a given feature of interest. (a and b) Effect of kVp on image quality, illustrating improved soft-tissue contrast at 60 kVp but fairly complex trade-offs in the underlying quantum noise. (c and d) Effect of pixel aperture on image quality (1×1 binning $= 0.194$ mm pixels, while 2×2 binning $= 0.388$ mm). The former favors bone detail and the latter soft-tissue discrimination. (e and f) Effect of reconstruction filter on image quality. Soft-tissue detection favors the smooth filter and bone detail the sharp filter, but the effect appears to be weaker than pixel binning—consistent with the d' calculations in Fig. 5.

difficult to interpret; (2) a global optimization is sure to yield different results depending on the imaging task, some of which may be intractable (e.g., separate optima in system magnification for separate tasks); and (3) with regard to parameters that impart a monotonic trend on d' (e.g., focal spot size, dose, and additive electronics noise), an optimization is unnecessary—e.g., concluding simply that optimal $a_{spot} = 0$ mm and optimal $\sigma_{add} = 0$ e. The observer models invoked in the detectability index of Eqs. (6)–(10) are perhaps the simplest form of prewhitening, nonprewhitening, match-filter type of ideal observers one can envision in linking fundamental Fourier metrics (NEQ) to imaging task. Further, 2D slice detectability models assume the slice to be infinitesimally thin, which is an approximation of real tomography. For thick slices, the slice detectability can be modified to include a transfer function (sinc) corresponding to the thickness of the slice, as in the "slice extraction" analysis of Siewerdsen *et al.*[61] Still, such metrics have shown reasonable correspondence with real observer performance,[10] and the brief, qualitative assessment of cadaver images acquired under similar conditions lend confidence that the trends suggested by the d' calculations are meaningful. This qualitative confirmation of trends in spatial resolution and contrast-to-noise ratio in cadaveric specimens falls short of a rigorous validation of the system optimization; however, a true optimization by physical experimentation is not only beyond the scope of the current manuscript (which aimed to provide a guide to system design) but also may be beyond the scope of practical measurement. Furthermore, analysis and optimization with respect to binary hypothesis-testing tasks does not pretend to convey the complexity of tasks performed by radiologists in reading diagnostic images, and the NPS, NEQ, and task-based detectability index should be considered as abstractions intended primarily as a guide to system design but not an absolute descriptor of diagnostic quality. Finally, the results considered three individual tasks with the potential for distinct optima. In cases where such optima concerned image processing or reconstruction parameters (e.g., the binning factor and choice of reconstruction filter), the results simply suggest parallel processing techniques. However, in cases, where such optima suggest distinct system design or image acquisition parameters (e.g., kVp or magnification), then a multitask optimization should be considered.

ACKNOWLEDGMENTS

The authors acknowledge countless valuable conversations with Dr. Ian Cunningham (Robarts Research Institute, London ON) regarding cascaded systems analysis, Dr. Angel Pineda (UC Fullerton, Fullerton CA) regarding noise stationarity and Fourier metrology, Dr. John Boone (UC Davis, Davis CA) on CBCT system design and the power-law nature of anatomical clutter, Dr. Norbert Pelc (Stanford University, Palo Alto CA) on NPS analysis and measurement, and Dr. Iacovos Kyprianou (FDA, Rockville MD) on NPS analysis and x-ray scatter. The work was supported by NIH 2R01-CA-112163 and academic-industry partnership with Carestream Health Inc. (Rochester NY).

APPENDIX A: MTF

The modulation transfer function (MTF) associated with different stages in the 3D cascaded systems model is summarized below. The MTF calculations account for detector cross-talk - e.g., spread of optical photons in the scintillator, integration over the pixel aperture, and any (believed to be small) charge sharing between photodiode elements. The MTF terms are defined in the 2D frequency domain (f_u, f_v) of the detector and related to the plane of the object (at isocenter) and the corresponding 3D frequency domain (f_x, f_y, f_z) by demagnification: $f_u = f/M$ and $f_v = f_z/M$, where M is the magnification and f is the radial frequency, $f = \sqrt{f_x^2 + f_y^2}$.

1. Focal spot MTF

The x-ray focal spot may be modeled as a uniform source distribution (e.g., rectangular aperture with characteristic width a_{spot}) as in Johns and Cunnigham.[62] The corresponding focal spot MTF in the detector plane depends on geometric magnification

$$T_{spot}(f_u, f_v) = \mathrm{sinc}\big((M-1)a_{spot}f_u\big)\,\mathrm{sinc}\big((M-1)a_{spot}f_v\big). \tag{A1}$$

The focal spot MTF degrades as the detector moves farther from isocenter (i.e., M increases) and as focal spot size (a_{spot}) increases.

2. Scintillator MTF

T_3 describes the stochastic spread of optical photons in the CsI:Tl scintillator and was approximated by a Lorentzian fit to empirical data[3]

$$T_3(f_u, f_v) = a.e^{-\left(\sqrt{f_u^2+f_v^2}\right)/b} + (1-a)\frac{1}{1+(f_u^2+f_v^2)/c^2}. \tag{A2}$$

In this work, the parameters were held fixed at $a = 0.27$, $b = 0.91$, and $c = 1.66$, corresponding to values measured for a 250 mg/cm^2 CsI:Tl converter as on the proposed extremities CBCT scanner. The dependence of H on kVp, and effects of oblique incidence were assumed negligible in the current work.

3. Photodiode aperture MTF

T_5 is the MTF associated with integration of quanta by the photodiode aperture

$$T_5(f_u, f_v) = |\mathrm{sinc}\,(f_u a_{pd})\,\mathrm{sinc}(f_v a_{pd})|. \tag{A3}$$

We assume a square photodiode, a_{pd}, with the fraction of the total pixel size (a_{pix}) that is sensitive to light given by $a_{pd} = a_{pix}\sqrt{f_{pd}}$, where f_{pd} is the pixel fill factor.

4. Apodization window

T_{11} represents the apodization (smoothing) window applied to the (ramp-filtered) projection data prior to back-projection—e.g., a cosine filter

$$T_{11}(f_u,f_v) = h_{\text{win}} + (1-h_{\text{win}})\cos 2\pi f_u a_{\text{pix}} \quad 0.5 \le h_{\text{win}} \le 1.0, \tag{A4}$$

where h_{win} is an adjustable smoothing parameter varying from 0.5 (Hann filter) to 1.0 (all-pass apodization giving a pure ramp (Ram-Lak) filter).

5. Bilinear interpolation

T_{12} describes bilinear interpolation (up-sampling) performed on the filtered projection to approximate a continuous image prior to the backprojection step

$$T_{12}(f_u,f_v) = \text{sinc}^2(f_u a_{\text{pix}})\,\text{sinc}^2(f_v a_{\text{pix}}). \tag{A5}$$

Other interpolation schemes, such as nearest-neighbor interpolation, may be used to up-sample the projection images, with MTF and potential aliasing effects described by Tward and Siewerdsen.[5]

APPENDIX B: X-RAY SCATTER

X-ray scatter is a significant source contrast degradation and artifacts in CBCT.[16] Muntz *et al.*[63] developed a scatter point-source model to describe the dependence of scatter-to-primary ratio (SPR) on system geometry. The model assumes scatter to originate from a virtual point source located at distance y_{scat} from the object plane (isocenter), typically in the direction of the primary x-ray source. Neitzel[64] employed the model and derived the selectivity of an air gap, Σ, leading to the following SPR formulation[16] in the context of CBCT geometry:

$$SPR = \frac{q_0^s}{q_0^p} = \frac{F'}{1-F'}\cdot\frac{1}{\Sigma} = \frac{F'}{1-F'}\cdot\frac{1}{\left(1-d_{\text{gap}}/SDD\right)^2\cdot\left(1+d_{\text{gap}}/y_{\text{scat}}\right)^2}, \tag{B1}$$

where F' is the scatter fraction at the exit of the object and is dependent on object size and field of view. In the results reported here, F' was held fixed at a value of 0.6, corresponding with low to moderate scatter conditions, as is the case for extremities.[16] Object plane to detector gap is given by $d_{\text{gap}} = SDD - SAD$ and y_{scat} is the effective scatter source location, taken to be fixed at 20 cm.[65] The SPR is seen to depend strongly on magnification. We assumed *SPR* to be independent of kVp, a simplifying assumption supported by Monte Carlo simulations in CBCT that show SPR to be a weak function of kVp in the diagnostic energy range.[66]

APPENDIX C: QUANTUM AND ELECTRONICS NOISE

1. Quantum noise

The propagation of the NPS in FPDs was modeled as a serial and parallel cascade of stages,[3,37] and the resulting 2D projection NPS (without additive electronic noise) is given by

$$S_7^{\text{tot}}(f_u,f_v) = \overline{q_0^{tot}}\,a_{pd}^4\bar{g}_1\bar{g}_2(1+\bar{g}_4 P_K T_3^2)T_5^2 * *III_6, \tag{C1}$$

where notation is consistent with previous work.[5] $\overline{q_0^{\text{tot}}}$ is the total incident fluence at the detector, $\bar{g}_1$ and $\bar{g}_2$ are the mean gains associated with interaction of x rays in the scintillator and conversion to secondary quanta, respectively, T_3 describes the spread of optical quanta, and P_K is a composite term associated with the parallel cascade arising from K-fluorescence as described by Yao and Cunningham;[34] $\bar{g}_4$ describes the coupling of optical photons to the photodiode aperture (a_{pd}), T_5 is the pixel aperture MTF, and III_6 represents the sampling of the image at discrete pixel locations.

The 2D NPS formulation has been extended to the 3D reconstruction cascade,[5,6] including postacquisition binning (T_8) and sampling (III_8), the log normalization step, ramp ($\bar{f}/M$) and apodization filter (T_{11}), interpolation (T_{12}), filtered backprojection ($1/f$), 3D sampling (III_{14}) stages and optional postreconstruction binning (T_{15}) and sampling (III_{15}). The resulting quantum noise-only 3D NPS is

$$S_Q^{\text{tot}}(f_x,f_v,f_z) = \left[\frac{\left((S_7^p+S_7^s)T_8^2\right) * *III_8}{\left(a_{pd}^2\left[\overline{q_0^p g_1^p g_2^p g_4^p}+\overline{q_0^s g_1^s g_2^s g_4^s}\right]\right)^2}\times\left(\frac{\bar{f}}{M}\right)^2 T_{11}^2 T_{12}^2\frac{1}{M^2}\frac{\pi M^2}{m}\frac{1}{f} * * * III_{14}\right] T_{15}^2 * * * III_{15}, \tag{C2}$$

where the superscript on S_Q^{p+s} denotes quantum NPS due to primary ($\overline{q_0^p}$) and scattered ($\overline{q_0^s}$) components of total incident fluence ($\overline{q_0^{\text{tot}}}$). The total, primary, and scatter fluence are related as: $(\overline{q_0^{\text{tot}}}) = (\overline{q_0^p}) + (\overline{q_0^s})$ and $SPR = \overline{q_0^s}/\overline{q_0^p}$. The above NPS formulation requires separate computation of projection NPS and gain terms for primary ($S_7^p, \overline{g_1^p}, \overline{g_2^p}$, etc.) and scattered ($S_7^s, \overline{g_1^s}, \overline{g_2^s}$, etc.) quanta. The scatter dependence of quantum NPS has been shown to be small. Also, the maximum energy shift associated with incoherent scatter is small. (For example, for a monoenergetic x-ray $E = 60$ keV scattered at $\theta = 45°$, the scattered x-ray has energy $E' = E - [E^2/(m_e c^2)](1-\cos\theta) = 57.9$ keV). We, therefore, make a simplifying assumption that the energy shift between primary and scattered spectra is negligible. In that case, the scatter NPS (S_7^s) takes the same form as S_7^p (except for the fluence term) and the 3D NPS reduces to

$$S_Q^{\text{tot}}(f_x,f_y,f_z) = \left[\frac{\left(\left(\overline{q_0^{\text{tot}}}\,a_{pd}^4\bar{g}_1\bar{g}_2(1+\bar{g}_4 P_K T_3^2)T_5^2 * *III_6\right)T_8^2\right) * *III_8}{\left(\overline{q_0^{\text{tot}}}\,a_{pd}^2\bar{g}_1\bar{g}_2\bar{g}_4\right)^2}\times\left(\frac{\bar{f}}{M}\right)^2 T_{11}^2 T_{12}^2\frac{1}{M^2}\frac{\pi M^2}{m}\frac{1}{f} * * * III_{14}\right] T_{15}^2 * * * III_{15}. \tag{C3}$$

2. Additive electronics noise

S_{add} represents the additive electronic noise associated with dark noise from the a-Si:H pixel elements, uncorrected voltage fluctuations on the gate and bias lines, amplifier

noise, and digitization noise in analog-to-digital converters.[3] The term is approximated as: $S_{\text{add}} = \left(\sigma_{\text{add}}.a_{\text{pix}}\right)^2$, where σ_{add} is the pixel variance measured in "dark-only" images. The additive NPS after propagating through the 3D reconstruction cascade is

$$S_E^{p+s}(f_x,f_y,f_z) = \left[\frac{(S_{\text{add}}T_8^2) * *III_8}{\left(a_{pd}^2\left[\overline{q_0^p g_1^p g_2^p g_4^p} + \overline{q_0^s g_1^s g_2^s g_4^s}\right]\right)^2} \times \left(\frac{\bar{f}}{M}\right)^2 T_{11}^2 T_{12}^2 \frac{1}{M^2}\frac{\pi M^2}{m}\frac{1}{f} * * * III_{14}\right] T_{15}^2 * * * III_{15}, \quad \text{(C4)}$$

where superscript on S_0^{p+s} denotes dependence on primary and scattered components of incident fluence and remaining terms are as described above. We invoke the same assumption as above: the energy shift between primary and scattered spectra is small, giving $\overline{g_1^p} = \overline{g_1^s}$, etc., and a simplified additive electronic NPS

$$S_E^{\text{tot}}(f_x,f_y,f_z) = \left[\frac{(S_{\text{add}}T_8^2) * *III_8}{\left(\overline{q_0^{\text{tot}}} a_{pd}^2 \bar{g}_1 \bar{g}_2 \bar{g}_4\right)^2}\left(\frac{\bar{f}}{M}\right)^2 T_{11}^2 T_{12}^2 \frac{1}{M^2}\frac{\pi M^2}{m}\frac{1}{f} * * * III_{14}\right] \times T_{15}^2 * * * III_{15}. \quad \text{(C5)}$$

APPENDIX D: ANATOMICAL BACKGROUND NOISE

The anatomical background power spectrum describes the fluctuations intrinsic to the object (clutter) and is frequently characterized by a power-law relationship used to describe a wide variety of random and semirandom processes[21,22]

$$S_B(f_{3D}) = \kappa/f_{3D}^{\beta}, \quad \text{(D1)}$$

where κ describes the magnitude of fluctuation, β is the degree of correlation of the variations, and f_{3D} is the 3D radial frequency, $f_{3D} = \sqrt{f_x^2 + f_y^2 + f_z^2}$. For calculations pertaining to the musculoskeletal scanner described above, the numerator κ was computed for a muscle vs water contrast background scaled from the experimentally determined value (κ_{acrylic}) measured in an acrylic sphere "clutter phantom" as follows:[30]

$$\kappa(\text{kVp}, SPR) = \kappa_{\text{acrylic}}\left(\frac{\frac{1}{1+SPR}\left(\mu_{\text{Muscle}}^{\text{eff}}(kVp) - \mu_{\text{Water}}^{\text{eff}}(\text{kVp})\right)}{\mu_{\text{Acrylic}}^{\text{eff}}}\right)^2, \quad \text{(D2)}$$

where $\kappa_{\text{acrylic}} = 3.72 \times 10^{-7}\ \mu^2\ \text{mm}^3$. The exponent $\beta = 3$ was calculated based on fractal properties of self-similar objects (such as equal volumes of different sized spheres in the generic phantom above) and is similar to widely reported values.[25–27,43] The numerator κ carries a dependence on kVp because of the effective attenuation coefficients of object and background, and decreases with scatter as $(1/1+SPR)^2$. κ and hence, anatomical clutter, decrease as a function of increasing kVp (reduced relative contrast). For instance, reduction in overlying anatomical clutter is a principal motivation for the use of higher kVp techniques in projection chest radiography.[36]

[a]Author to whom correspondence should be addressed. Electronic mail: jeff.siewerdsen@jhu.edu; Telephone: 443-287-6269; Fax: 410-955-9826.

[1]W. Zbijewski, P. D. Jean, P. Prakash, Y. Ding, J. W. Stayman, N. Packard, R. Senn, D. Yang, J. Yorkston, A. Machado, J. A. Carrino, and J. H. Siewerdsen, "Design and optimization of a dedicated cone-beam CT system for musculoskeletal extremities imaging," Proc. SPIE, **7961**, 796104 (2011).
[2]W. Zbijewski, P. D. Jean, P. Prakash, Y. Ding, J. W. Stayman, N. Packard, R. Senn, D. Yang, J. Yorkston, A. Machado, J. A. Carrino, and J. H. Siewerdsen, "A dedicated cone-beam CT system for musculoskeletal extremities imaging: Design, optimization, and initial performance characterization," Med. Phys. **38**(8), 4700–4713 (2011).
[3]J. H. Siewerdsen, L. E. Antonuk, Y. el-Mohri, J. Yorkston, W. Huang, J. M. Boudry, and I. A. Cunningham, "Empirical and theoretical investigation of the noise performance of indirect detection, active matrix flat-panel imagers (AMFPIs) for diagnostic radiology," Med. Phys. **24**, 71–89 (1997).
[4]W. Zhao and J. A. Rowlands, "Digital radiology using active matrix readout of amorphous selenium: Theoretical analysis of detective quantum efficiency," Med. Phys. **24**, 1819–1833 (1997).
[5]D. J. Tward and J. H. Siewerdsen, "Cascaded systems analysis of the 3D noise transfer characteristics of flat-panel cone-beam CT," Med. Phys. **35**, 5510–5529 (2008).
[6]D. J. Tward and J. H. Siewerdsen, "Noise aliasing and the 3D NEQ of flat-panel cone-beam CT: Effect of 2D/3D apertures and sampling," Med. Phys. **36**, 3830–3843 (2009).
[7]A. Ganguly, S. Rudin, D. R. Bednarek, and K. R. Hoffmann, "Micro-angiography for neuro-vascular imaging. II. Cascade model analysis," Med, Phys, **30**, 3029–3039 (2003).
[8]S. Vedantham, A. Karellas, and S. Suryanarayanan, "Solid-state fluoroscopic imager for high-resolution angiography: Parallel-cascaded linear systems analysis," Med. Phys. **31**, 1258–1268 (2004).
[9]International Commission on Radiation Units and Measurements, "Medical imaging—the assessment of image quality," ICRU Report No. 54, (International Commission on Radiation Units and Measurements, Bethesda, MD, 1996).
[10]G. J. Gang, J. Lee, J. W. Stayman, D. J. Tward, W. Zbijewski, J. L. Prince, and J. H. Siewerdsen, "Analysis of Fourier-domain task-based detectability index in tomosynthesis and cone-beam CT in relation to human observer performance," Med. Phys. **38**, 1754–1768 (2011).
[11]J. H. Siewerdsen and D. A. Jaffray, "Cone-beam CT with a flat-panel imager: Noise considerations for fully 3D computed tomography," Proc. SPIE, **3977**, 408–416 (2000).
[12]J. H. Siewerdsen and D. A. Jaffray, "Cone-beam computed tomography with a flat-panel imager: Magnitude and effects of x-ray scatter," Med. Phys. **28**, 220–231 (2001).
[13]J. H. Siewerdsen and D. A. Jaffray, "Three-dimensional NEQ transfer characteristics of volume CT using direct- and indirect-detection flat-panel imagers," Proc. SPIE, **5030**, 92–102 (2003).
[14]S. Yoon, J. G. Gang, D. J. Tward, J. H. Siewerdsen, and R. Fahrig, "Analysis of lung nodule detectability and anatomical clutter in tomosynthesis imaging of the chest," Proc. SPIE, **7258**, 72581M (2009).
[15]B. Zhao, J. Zhou, Y.-H. Hu, T. Mertelmeier, J. Ludwig, and W. Zhao, "Experimental validation of a three-dimensional linear system model for breast tomosynthesis," Med. Phys. **36**, 240–251 (2009).
[16]J. H. Siewerdsen and D. A. Jaffray, "Optimization of x-ray imaging geometry (with specific application to flat-panel cone-beam computed tomography)," Med Phys **27**, 1903–1914 (2000).
[17]A. Jain, A. T. Kuhls-Gilcrist, S. K. Gupta, D. R. Bednarek, and S. Rudin, "Generalized two-dimensional (2D) linear system analysis metrics (GMTF, GDQE) for digital radiography systems including the effect of focal spot, magnification, scatter, and detector characteristics," Proc. Soc. Photo-Opt. Instrum. Eng. **7622** (2010).
[18]I. S. Kyprianou, S. Rudin, D. R. Bednarek, and K. R. Hoffmann, "Generalizing the MTF and DQE to include x-ray scatter and focal spot

unsharpness: Application to a new microangiographic system," Med. Phys. **32**, 613–626 (2005).
[19]A. Ganguly, S. Rudin, D. R. Bednarek, K. R. Hoffmann, and I. S. Kyprianou, "Micro-angiography for neuro-vascular imaging. I. Experimental evaluation and feasibility," Med. Phys. **30**, 3018–3028 (2003).
[20]I. S. Kyprianou, S. Rudin, D. R. Bednarek, and K. R. Hoffmann, "Study of the generalized MTF and DQE for a new microangiographic system," Proc. SPIE **5368**, 349–360 (2004).
[21]F. O. Bochud, C. K. Abbey, and M. P. Eckstein, "Visual signal detection in structured backgrounds. III. Calculation of figures of merit for model observers in statistically nonstationary backgrounds," J. Opt. Soc. Am. A Opt. Image Sci. Vis. **17**, 193–205 (2000).
[22]A. E. Burgess, "Mammographic structure: Data preparation and spatial statistics analysis," Proc. SPIE, **3661**, 642–653 (1999).
[23]A. E. Burgess, F. L. Jacobson, and P. F. Judy, "Human observer detection experiments with mammograms and power-law noise," Med. Phys. **28**, 419–437 (2001).
[24]J. J. Heine, S. R. Deans, R. P. Velthuizen, and L. P. Clarke, "On the statistical nature of mammograms," Med. Phys. **26**, 2254–2265 (1999).
[25]K. G. Metheany, C. K. Abbey, N. Packard, and J. M. Boone, "Characterizing anatomical variability in breast CT images," Med. Phys. **35**, 4685–4694 (2008).
[26]E. Engstrom, I. Reiser, and R. Nishikawa, "Comparison of power spectra for tomosynthesis projections and reconstructed images," Med. Phys. **36**, 1753–1758 (2009).
[27]S. J. Glick, S. Vedantham, and A. Karellas, "Investigation of optimal kVp settings for CT mammography using a flat-panel imager," Proc. SPIE, **4682**, 392–402 (2002).
[28]X. Gong, S. J. Glick, B. Liu, A. A. Vedula, and S. Thacker, "A computer simulation study comparing lesion detection accuracy with digital mammography, breast tomosynthesis, and cone-beam CT breast imaging," Med. Phys. **33**, 1041–1052 (2006).
[29]I. Reiser and R. M. Nishikawa, "Task-based assessment of breast tomosynthesis: Effect of acquisition parameters and quantum noise," Med. Phys. **37**, 1591–1600 (2010).
[30]G. J. Gang, D. J. Tward, J. Lee, and J. H. Siewerdsen, "Anatomical background and generalized detectability in tomosynthesis and cone-beam CT," Med. Phys. **37**, 1948–1965 (2010).
[31]J. H. Siewerdsen, D. J. Moseley, and D. A. Jaffray, "Incorporation of task in 3D imaging performance evaluation: The impact of asymmetric NPS on detectability," Proc. SPIE, **5368**, 89–97 (2004).
[32]E. Samei, M. J. Flynn, and W. R. Eyler, "Detection of subtle lung nodules: Relative influence of quantum and anatomic noise on chest radiographs," Radiology **213**, 727–734 (1999).
[33]I. A. Cunningham, M. S. Westmore, and A. Fenster, "A Spatial-frequency dependent quantum accounting diagram and detective quantum efficiency model of signal and noise-propagation in cascaded imaging-systems," Med. Phys. **21**, 417–427 (1994).
[34]J. Yao and I. A. Cunningham, "Parallel cascades: New ways to describe noise transfer in medical imaging systems," Med. Phys. **28**, 2020–2038 (2001).
[35]S. Vedantham, A. Karellas, S. Suryanarayanan, D. Albagli, S. Han, E. J. Tkaczyk, C. E. Landberg, B. Opsahl-Ong, P. R. Granfors, I. Levis, C. J. D'Orsi, and R. E. Hendrick, "Full breast digital mammography with an amorphous silicon-based flat panel detector: Physical characteristics of a clinical prototype," Med. Phys. **27**, 558–567 (2000).
[36]S. Richard, J. H. Siewerdsen, D. A. Jaffray, D. J. Moseley, and B. Bakhtiar, "Generalized DQE analysis of radiographic and dual-energy imaging using flat-panel detectors," Med. Phys. **32**, 1397–1413 (2005).
[37]S. Richard and J. H. Siewerdsen, "Optimization of dual-energy imaging systems using generalized NEQ and imaging task," Med. Phys. **34**, 127–139 (2007).
[38]I. A. Cunningham, "Analyzing system performance," *The Expanding Role of Medical Physics in Diagnostic Imaging*, edited by G. D. Frey and P. Sprawls (AAPM, Madison, WI, 1997), pp. 231–263.
[39]I. A. Cunningham and R. Shaw, "Signal-to-noise optimization of medical imaging systems," J. Opt. Soc. Am. A **16**, 621–632 (1999).
[40]H. H. Barrett, "Objective assessment of image quality: Effects of quantum noise and object variability," J. Opt. Soc. Am. A **7**, 1266–1278 (1990).
[41]H. H. Barrett, J. L. Denny, H. C. Gifford, C. K. Abbey, R. F. Wagner, and K. J. Myers, "Generalized NEQ: Fourier analysis where you would least expect to find it," Proc. SPIE, **2708**, 41–52 (1996).
[42]J. A. Seibert and J. M. Boone, "X-ray scatter removal by deconvolution," Med. Phys. **15**, 567–575 (1988).
[43]R. M. Nishikawa and M. J. Yaffe, "Signal-to-noise properties of mammographic film-screen systems," Med. Phys. **12**, 32–39 (1985).
[44]J. H. Siewerdsen and L. E. Antonuk, "DQE and system optimization for indirect-detection flat-panel imagers in diagnostic radiology," Proc. SPIE, **3336**, 546–555 (1998).
[45]R. F. Wagner, D. G. Brown, and M. S. Pastel, "Application of information theory to the assessment of computed tomography," Med. Phys. **6**, 83–94 (1979).
[46]K. J. Myers, H. H. Barrett, M. C. Borgstrom, D. D. Patton, and G. W. Seeley, "Effect of noise correlation on detectability of disk signals in medical imaging," J. Opt. Soc. Am. A **2**, 1752–1759 (1985).
[47]A. E. Burgess, X. Li, and C. K. Abbey, "Visual signal detectability with two noise components: Anomalous masking effects," J. Opt. Soc. Am. A Opt. Image Sci. Vis. **14**, 2420–2442 (1997).
[48]P. G. J. Barten, *Contrast Sensitivity of the Human Eye and Its Effects on Image Quality* (SPIE, Bellingham, 1999).
[49]S. Richard and J. H. Siewerdsen, "Comparison of model and human observer performance for detection and discrimination tasks using dual-energy x-ray images," Med. Phys. **35**, 5043–5053 (2008).
[50]J. H. Siewerdsen, A. M. Waese, D. J. Moseley, S. Richard, and D. A. Jaffray, "Spektr: A computational tool for x-ray spectral analysis and imaging system optimization," Med. Phys. **31**, 3057–3067 (2004).
[51]J. M. Boone and J. A. Seibert, "An accurate method for computer-generating tungsten anode x-ray spectra from 30 to 140 kV," Med. Phys. **24**, 1661–1670 (1997).
[52]J. H. Hubbell and S. M. Seltzer, "Tables of x-ray mass attenuation coefficients and mass energy-absorption coefficients," NIST Report No. 5632 (1995).
[53]International Commission on Radiation Units and Measurements, "Photon, electron, proton and neutron interaction data for body tissues," ICRU Report No. 46 (International Commission on Radiation Units and Measurements, Bethesda, MD, 1992).
[54]I. A. Cunningham, M. Sattarivand, G. Hajdok, and J. Yao, "Can a Fourier-based cascaded-systems analysis describe noise in complex shift-variant spatially sampled detectors?," Proc. SPIE, **5368**, 79–88 (2004).
[55]M. Albert and A. D. A. Maidment, "Linear response theory for detectors consisting of discrete arrays," Med. Phys. **27**, 2417–2434 (2000).
[56]E. Samei, J. T. Dobbins, III, J. Y. Lo, and M. P. Tornai, "A framework for optimising the radiographic technique in digital x-ray imaging," Radiat. Prot. Dosim. **114**, 220–229 (2005).
[57]R. L. McKinley, E. Samei, C. N. Brzymialkiewicz, M. P. Tornai, and J. Carey E. Floyd, *Measurements of an Optimized Beam for X-Ray Computed Mammotomography* (SPIE, n, 2004).
[58]N. A. Shkumat, J. H. Siewerdsen, A. C. Dhanantwari, D. B. Williams, S. Richard, N. S. Paul, J. Yorkston, and R. Metter, "Optimization of image acquisition techniques for dual-energy imaging of the chest," Med. Phys. **34**, 3904–3915 (2007).
[59]L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," J. Opt. Soc. Am. A **1**, 612–619 (1984).
[60]D. Biswas, J. E. Bible, M. Bohan, A. K. Simpson, P. G. Whang, and J. N. Grauer, "Radiation exposure from musculoskeletal computerized tomographic scans," J. Bone Jt. Surg., Am. **91**, 1882–1889 (2009).
[61]J. H. Siewerdsen, I. A. Cunningham, and D. A. Jaffray, "A framework for noise-power spectrum analysis of multidimensional images," Med. Phys. **29**, 2655–2671 (2002).
[62]H. E. Johns and J. R. Cunningham, *The Physics of Radiology*, 4th ed. (Thomas, Springfield, IL, 1983).
[63]E. P. Muntz, G. Jacobson, E. M. Kaegi, and D. J. Klein, "Electronic grids for electrostatic imaging systems," Radiology **121**, 197–204 (1976).
[64]U. Neitzel, "Grids or air gaps for scatter reduction in digital radiography: A model calculation," Med. Phys. **19**, 475–481 (1992).
[65]J. A. Sorenson and J. Floch, "Scatter rejection by air gaps: An empirical model," Med. Phys. **12**, 308–316 (1985).
[66]Y. Kyriakou and W. A. Kalender, "X-ray scatter data for flat-panel detector CT," Phys. Med. **23**, 3–15 (2007).